METABOLISM OF GUT MICROBIOTA REGULATES EPIGENETIC EVENTS IN HOST TISSUES

Joan Miró Blanch

Joan Miró Blanch

Metabolism of gut microbiota regulates epigenetic events in host tissues

PhD Thesis Dissertation

Supervised by:
Dr. Óscar Yanes Torrado

Department of Medicine and Surgery

This dissertation is submitted for the degree of Doctor of Philosophy

UNIVERSITAT ROVIRA i VIRGILI

Reus 2023

UNIVERSITAT ROVIRA i VIRGILI

FAIG CONSTAR que aquest treball, titulat "Metabolism of gut microbiota regulates epigenetic events in host tissues", que presenta Joan Miró Blanch per a l'obtenció del títol de Doctor, ha estat realitzat sota la meva direcció al Departament Medicina i Cirurgia d'aquesta universitat.

HAGO CONSTAR que el presente trabajo, titulado "Metabolism of gut microbiota regulates epigenetic events in host tissues", que presenta Joan Miró Blanch para la obtención del título de Doctor, ha sido realizado bajo mi dirección en el Departamento Medicina y Cirurgia de esta universidad.

I STATE that the present study, entitled "Metabolism of gut microbiota regulates epigenetic events in host tissues", presented by Joan Miró Blanch for the award of the degree of Doctor, has been carried out under my supervision at the Department Medicine and surgery of this university.

[Ciutat], [data] / [Ciudad], [fecha] / [City], [date]

El/s director/s de la tesi doctoral
El/los director/es de la tesis doctoral
Doctoral Thesis Supervisor/s

OSCAR YANES TORRADO - DNI 45539680W

Digitally signed by OSCAR YANES TORRADO - DNI 45539680W
Date: 2023.05.29 08:56:57 +02'

[signatura] / [firma] / [signature]

[nom / [nombre] / [name]

[signatura] / [firma] / [signature]

[nom] / [nombre] / [name]

Acknowledgments i agraïments

Al meu supervisor Oscar Yanes, per la teva amplia visió, per guiar-me i recolzar-me en l'aventura de la ciència, i haver-me donat l'oportunitat de fer recerca en el meu "dream project".
Al Jordi, per ensenyar-me tantes i tantes coses sobre R i metabolomica, i haver compartit llargues discussions sobre el meu i altres projectes, però sobretot per les teves frases mítiques i amistat.
A la Sandra, per acompanyar-me en el fascinant món de la metabolómica experimental i la espectrometria de masses, i per que al final si que podem parlar i entendren's com amics.
Al Jordi R. per l'ajuda en automatitzar l'anàlisis de dades i continuar amb el projecte.
A tot el grup Mil@b, per haver compartit tants seminaris interessants i idees sobre la metabolomica.
A Pau y Pablo, por apoyar mi investigación con muestras y discusiones sobre ciencia.
A la Marta, per acollir-me al seu lab i donar-me l'oportunitat d'utilitzar el Marenostrum.
A Teresa, Salva y Ana, por acogerme en Valencia y enseñarme estadística multivariante.
To Axel, for hosting me in his house and his lab in Munich, and teaching me how histone's work.
To Ignasi, Magdalini, Anuroop and Andreas, for teaching me the pitfalls of preparing histone digest and for facilitating and enhancing the internship in Munich.
To all ChroMe fellows, for sharing the first years of the PhD with wonderful scientific and life experiences, and for your life-long friendships.
Al Bo, per marcar-me tant durant Biologia i per sempre estar disposat a escoltar i parlar de ciencia.
Als meus amics, per distreure'm, escalar, alegrar-me i recolzar-me durant aquests anys.
Al meu pare, per saber mantenir la distància justa i pel teu amor.
A la meva mare, per la transformació que has fet a la teua vida i pel teu amor.
Al meu germà, per mostrar-me sempre una nova perspectiva, com ser millor persona i pel teu amor.
Als meus 3 fills, per la puresa de les vostres accions i l'amor incondicional que em mostreu cada dia.
A toi Tania, pour toute la patience et le soutien que tu m'as montré depuis le début, aussi a la smala, merci pour m'accompagner dans la vie et l'aventure d'être parents, et pour ton amour. Aquesta tesis está dedicada a tú.

To Tania

for her explicit love,
her latent support,
her soft energy,
and her hidden strength.

"Life did not take over the world by combat, but by networking."
-Lynn Margulis

"The time has come to replace the purely reductionist 'eyes-down' molecular perspective with a new and genuinely holistic, eyes-up, view of the living world, one whose primary focus is on evolution, emergence, and biology's innate complexity."
-Carl Woese

UNIVERSITAT ROVIRA I VIRGILI
METABOLISM OF GUT MICROBIOTA REGULATES EPIGENETIC EVENTS IN HOST TISSUES
Joan Miró Blanch

Table of contents

List of tables and figures

Introduction tables and figures

Chapter 1 tables and figures

Chapter 2 tables and figures

Chapter 3 tables and figures

List of publications

Miro-Blanch J and Yanes O (2019) Epigenetic Regulation at the Interplay Between Gut Microbiota and Host Metabolism. *Front. Genet*. 10:638. doi: 10.3389/fgene.2019.00638

J. Miro-Blanch, A. Junza, J. Capellades, A. Balvay, C. Maudet, S. Raineri, M. Kovatcheva, S. Rabot, J. Mellor, M. Serrano, O. Yanes. A targeted metabolomics method to detect epigenetically relevant metabolites. Journal of Biological Chemistry (Submitted, 2023)

J. Miro-Blanch, J. Capellades, J. Rofes, A. Junza, C. Heyer, I. Forner, M. Serefidou, L. Santus, M. B. Carbonetto, T. García-Barrera, T. Rubio, V. Chapaprieta, A. Balvay, C. Maudet, S. Rabot, J.I Martín-Subero, A Conesa, P. González-Torres, T. Gabaldon, A. Imhof, M. Melé, M. Schlesner and O. Yanes. Gut *Ruminococcaceae* are associated with hypermethylation of genes controlling testosterone degradation in healthy mice males. BMJ Gut or Cell Host & Microbe (In preparation, 2023)

J. Miro-Blanch, J. Capellades, J. Rofes, A. Junza, C. Heyer, L. Santus, I. Forner, M. Serefidou, M. B. Carbonetto, T. García-Barrera, T. Rubio, V. Chapaprieta, A. Balvay, C. Maudet, S. Rabot, J.I Martín-Subero, A Conesa, P. González-Torres, A. Imhof, M. Melé, M. Schlesner and O. Yanes. Gut Microbiome and host metabolome, transcriptome, metalome and epigenome measurements in a germ-free mouse model. Scientific Data (In preparation, 2024)

J. Miro-Blanch, J. Rofes, C. Heyer, P Gama, J. Capellades, A. Junza, L. Santus, I. Forner, M. B. Carbonetto, T. García-Barrera, T. Rubio, V. Chapaprieta, A. Balvay, C. Maudet, S. Rabot, J.I Martín-Subero, A Conesa, P. González-Torres, T. Gabaldon, A. Imhof, M. Melé, P. Garcia-Roves, M. Schlesner and O. Yanes. Lifestyle intervention alters gut microbiota and epigenetic marks in the host. BMJ Gut or Cell Host Microbe (In preparation, 2024)

J. Miro-Blanch, J. Rofes, C. Heyer, P. Gama, J. Capellades, A. Junza, C. Heyer, L. Santus, I. Forner, M. B. Carbonetto, T. García-Barrera, T. Rubio, V. Chapaprieta, A. Balvay, C. Maudet, S. Rabot, J.I Martín-Subero, A Conesa, P. González-Torres, T. Gabaldon A. Imhof, M. Melé, P. Garcia-Roves, M. Schlesner and O. Yanes. Microbiome and host metabolome, transcriptome, metalome and epigenome measurements in a male lifestyle intervention mouse model. Scientific Data (In preparation, 2024)

Contributions to other publications

M. Valera-Alberni, M. Joffraud, **J. Miro-Blanch**, J. Capellades, A. Junza, L. Dayon, A. Núñez Galindo, J. L. Sanchez-Garcia, A. Valsesia, A. Cercillieux, F. Söllner, A. G. Ladurner, O. Yanes, C. Cantó, Crosstalk between Drp1 phosphorylation sites during mitochondrial remodeling and their impact on metabolic adaptation. *Cell Rep.* **36**, 109565 (2021).

I. Huber-Ruano, E. Calvo, J. Mayneris-Perxachs, M.-M. Rodríguez-Peña, V. Ceperuelo-Mallafré, L. Cedó, C. Núñez-Roa, **J. Miro-Blanch**, M. Arnoriaga-Rodríguez, A. Balvay, C. Maudet, P. García-Roves, O. Yanes, S. Rabot, G. M. Grimaud, A. De Prisco, A. Amoruso, J. M. Fernández-Real, J. Vendrell, S. Fernández-Veledo, Orally administered Odoribacter laneus improves glucose control and inflammatory profile in obese mice by depleting circulating succinate. *Microbiome*. **10**, 135 (2022)

A. Gonzalez-Franquesa, P. Gama-Perez, M. Kulis, K. Szczepanowska, N. Dahdah, S. Moreno-Gomez, A. Latorre-Pellicer, R. Fernández-Ruiz, A. Aguilar-Mogas, A. Hoffman, E. Monelli, S. Samino, **J. Miró-Blanch**, G. Oemer, X. Duran, E. Sanchez-Rebordelo, M. Schneeberger, M. Obach, J. Montane, G. Castellano, V. Chapaprieta, W. Sun, L. Navarro, I. Prieto, C. Castaño, A. Novials, R. Gomis, M. Monsalve, M. Claret, M. Graupera, G. Soria, C. Wolfrum, J. Vendrell, S. Fernández-Veledo, J. A. Enríquez, A. Carracedo, J. C. Perales, R. Nogueiras, L. Herrero, A. Trifunovic, M. A. Keller, O. Yanes, M. Sales-Pardo, R. Guimerà, M. Blüher, J. I. Martín-Subero, P. M. Garcia-Roves, Remission of obesity and insulin resistance is not sufficient to restore mitochondrial homeostasis in visceral adipose tissue. *Redox Biol*. **54**, 102353 (2022)

Abstract

Title
Metabolism of gut microbiota regulates epigenetic events in host tissues

Name
Joan Miró Blanch

Summary
How the gut microbiota and its host communicate and react together to environmental stimuli, is essential to understand a mammalian holobiont. In a biomedical context, such communications might be a reason for staying healthy if it is balanced, or associated with disease when the communication is disrupted. We propose here the "nutrient-microbe metabolism-host epigenetic axis", as a poorly understood axis of communication between the host and its microbiota. Microbiota and host tissues produce metabolites that can regulate and limit epigenetic enzymes, controlling epigenetic marks in the host. However, measuring the physico-chemical variety of epigenetically relevant metabolites in a single metabolomic method has been historically challenging. To overcome this limitation, we have created and optimized a targeted metabolomics method to measure over 30 epigenetically relevant metabolites including acetyl-CoA, S-adenosylmethionine (SAM) or short chain fatty acids (SCFAs), at the same time and in the same sample extract.
Next, we used a germ-free mouse model, including males and females, to investigate how biological sex and microbiota status influenced the "nutrient-microbe metabolism-host epigenetic" axis of communication. We studied microbes, metabolism, epigenetic marks and gene expression in the host, focusing on the liver as target tissue, through a multi-omics and correlational analysis approach. The results of this thesis show the interaction of sex and microbiota in controlling the levels of DNA methylation in an additive manner. In particular, we have observed the additive effect of sex and microbiota in the levels of DNA methylation, where males are hypomethylated compared to females, and the microbiota accentuates this pattern in both sexes, being more pronounced in males. Males without microbiota present a feminized pattern. Additionally, we have observed a strong positive correlation between the levels of DNA methylation in genes involved in testosterone degradation, and the *Ruminococcaceae* family, more abundant in males than females. Altogether, we propose *Ruminococcaceae* as a key ecological player in the holobiont ecosystem, influencing host gene regulation in a sex-dependent manner, particularly through the possible modulation of methylation levels of testosterone degradation genes, necessary for steroid metabolism.

Resum

Títol
El metabolisme de la microbiota regula les marques epigenètiques en teixits de l'hoste

Nom
Joan Miró Blanch

Resum
Com la microbiota i el seu hoste es comuniquen i reaccionen a estímuls ambientals, és essencial per entendre un holobiont com un mamífer. En un context biomèdic, la seva comunicació és una de les raons per seguir amb salut si la comunicació es equilibrada, o esta associada a malalties si la comunicació es pertorba. Proposem aquí, l'axis "Nutrició-metabolisme microbià--epigenètica de l'hoste", com a axis de comunicació entre l'hoste i la seva microbiota poc estudiat. On la microbiota i l'hoste, produeixen metabòlits que poden regular i limitar enzims epigenètics, i controlar les marques epigenètics en l'hoste. No obstant, mesurar la varietat fisicoquímica d'aquest metabòlits epigenèticament rellevants amb un sol mètode metabolòmic ha estat històricament difícil. Per superar aquesta limitació, hem creat i optimitzat un mètode metabolomic dirigit que pot mesurar més de 30 metabòlits epigenèticament rellevants incloent l'acetyl-CoA, la S-adenosilmetionina o els àcids grassos de cadena curta, al mateix temps i en la mateixa extracció.
Després, hem utilitzat un model amb ratolins sense microbiota i amb microbiota, incloent mascles i femelles, per investigar com el sexe i la microbiota influencien l'axis de comunicació "Nutrició-metabolisme microbià--epigenètica de l'hoste". Estudiem els microbis, el metabolisme, les marques epigenètics i l'expressió de gens en l'hoste, i enfocant-nos en el fetge com a teixit diana, a través d'una aproximació multi-omica i un anàlisis correlacional. Els resultats d'aquesta tesis ensenyen la interacció del sexe i la microbiota de manera additiva en els nivells de metilació del ADN. Em observat aquesta additivitat entre el sexe i la microbiota, ja que els mascles estan hipometilats respecte les femelles, i la microbiota accentua aquest patró en els dos sexes, essent més pronunciat en mascles. Presentant els mascles sense microbiota un patró feminitzat. També hem observat una forta correlació positiva entre els nivells de metilació de l'ADN en gens involucrats en la degradació de la testosterona i la familia *Ruminococcaceae*, més abundant en mascles que en femelles. En conjunt, proposem els *Ruminococcaceae* com a peces clau en l'ecologia del holobiont, influenciant la regulació dels gens de l'hoste de manera diferent segons el sexe, via la modulació dels nivells de metilació en gens relacionats amb la degradació de la testosterona, necessaris pel metabolisme dels esterols.

Introduction

In the Introduction section, I describe the biological context and molecular layers involved in the host-microbiota communication, and why I embraced the holobiont theory to study this communication, in the first half. And introduce some technicalities about the set of omic technologies we used to study the different molecular layers in the holobiont, in the second half.

1. Influence of a microbial world

The biosphere, the complex network of ecosystems present on Earth, undoubtedly is a microbial dominated world, with microorganisms spreading and colonizing almost every aquatic, terrestrial and aerial environment possible, and present in, or on all life forms on this planet. We humans, we are no exception, we are covered with microorganisms on us and within us, in almost every organ of our bodies, in a similar cell ratio of human to bacterial cells (*1*). However, the number of genes encoded in our genomes are strikingly outnumbered by the number of genes encoded in our microbiome by a factor of at least 100 times more microbial genes, than human ones (*2*). These overwhelming higher numbers of genes present in our microbiome, arm microbes with a higher metabolic potential and functionality compared to the human genome. The microbial world plays key functions in human physiology, ecology, and evolution, and the ecological pressure exerted by these microbes has been also applied to many of our ancestors. From parents and grandparents, to the first hominids or the first mammalian organism, we all have been exposed and co-evolved with the microbes' influence. Meaning that, since the apparition of microbes on Earth, every evolutionary relationship has been shaped and intimately related to microbes.

This introductory section (1) aims to put you in an evolutionary mindset. I want you to understand that the research I conducted during my PhD thesis might have important evolutionary implications. Although evolution is not the main focus of my PhD and I do not enter this fascinating field much, I intend to give you here an evolutionary perspective on the main subject of the PhD. By guiding you through the evolution of life, and briefly describe how eukaryotes and, especially mammals, gain microscopic life-long companions, I hope you can better understand the main subject and results of my PhD research.

1.1. Evolution of life

Before life appeared on planet Earth, it was an anoxygenic, high temperature and reductive place where physic and chemistry reactions set the conditions for the spontaneous apparition of life (*3*) (Figure 1). Life originated on Earth, first with self-replicative organic molecules increasing in complexity over time (*4*). Second, a membrane-like system able to self-replicate, self-assemble and auto-catalize appeared and set the basis for the origin of the first living cell/organism (*5*), the **last universal common/cellular ancestor (LUCA)**. The first living cell/organism was a very simple and rudimentary living cell (*6*). However, going from this first single cell ancestor, to a multicellular organism such a mouse or a plant,

thousands of important evolutionary events had occurred, ultimately giving origin to the rich and diverse number of life forms on Earth.

Life appeared on Earth somewhere between 3.4-3.5 billion years ago, and *Bacteria* and *Archaea* appeared around a billion years later. *Archaea,* but especially *Bacteria* were, are and will be the most abundant and diverse organisms on planet Earth, thanks to their plasticity to adapt to a changing environment and evolving quickly. They play essential roles in the geological and physico-chemical cycles of organic and inorganic matter (*7*, *8*). They shape evolution, landscapes and life with a myriad of functions encoded in their genomes (*8*). But, *Bacteria* and *Archaea* were long present when the third domain of life appeared, the *Eukarya,* which will lately evolve to a diversity of organisms including mammals. Since *Bacteria* and *Archaea* were long present on Earth when *Eukarya* appeared, eukaryotes have developed very complex and intimate relationships with them (*9*). To the extreme that certain eukaryotes depend on them to survive, sense and adapt to their changing environment, or helping the host switch on and off genes. An is within these last important concepts where my research is relevant.

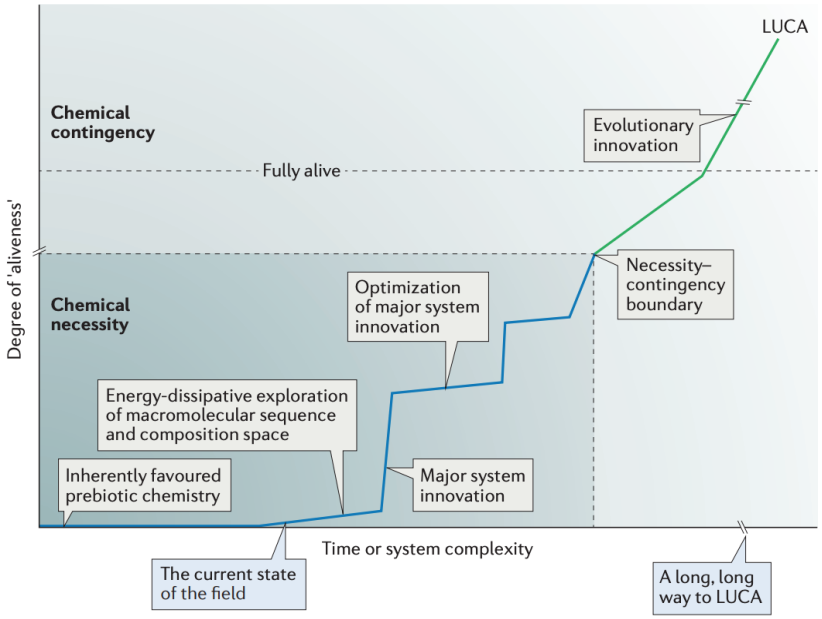


Figure 1. Origin of life. Schematic view of the transition from a "un-alive" chemical world to a more "alive" chemical world, with major evolutionary conditions and innovations to reach the last universal common ancestor (LUCA). From Sutherland (*5*).

Over millions of years, LUCA evolved from a rudimentary cell form, to a more complex cell organism, originating the first two domains of life that appeared on Earth: the *Bacteria* and the *Archaea*. They appeared somewhere between 2.5-3.2 billion years ago and were unicellular organisms without a nucleus. The third domain of life is estimated to appear somewhere between 1.7-2.1 billion years ago, the *Eukarya*, initially unicellular and later multicellular organisms with a true nucleus in their cells. To date, it is accepted that the *Eukarya* domain appeared on Earth later than the *Bacteria* and the *Archaea* ancestors, from a symbiotic relation between two organisms of these two domains. The *endosymbiont* theory coined by Lyn Margulis is widely accepted and postulates that an independent *Bacteria* and an

independent *Archaea* embarked in a mutualistic symbiosis relationship that over time became the *Eukarya* domain (*10*). A serie of engulfment events between an *Archaea* phylum called *Asgard* as the host cell with many eukaryotic features, and a *Bacteria* from the *Alphaproteobacteria* class as the endosymbiont cell that will later become the eukaryotic mitochondria, give rise to the *Eukarya* domain. The acquisition of the mitochondria is believed to be the reason to allow the eukaryotic cells to increase their complexity over time (*11*). This evolutionary (and revolutionary) innovation led to the diversification of the *Eukarya* domain, they evolved to a range of uni and pluricellular forms, and most of the macro-organisms that we know today form part of this domain. However, as observed in Figure 2, the tree of life is widely dominated by microorganisms belonging to the *Bacteria* domain first, and *Archaea* domains second, being *Eukarya* the less phylogenetic divers, and Bacteria, with a new proposed candidate phyla radiation, being the most diverse branch in the tree of life (*12*).

Another important event in the diversification of life as we know it today on planet Earth and after the apparition of the first microorganisms, is the rise of oxygen levels. The great oxygenation event (GOE) raised the oxygen levels from less than 0.001% to similar levels as we have today (*13*). The oxygen was produced continuously by microorganisms over a few million years, thanks to another evolutionary innovation, the oxygenic photosynthesis of *Cyanobacteria*. But, photosynthesis did not appear on earth only in an oxygenic mode, but with a diverse source of electron donors such as hydrogen, iron or hydrogen sulfide instead of water molecules (*14*). During and after the GOE, life on Earth has adapted to the increasing levels of oxygen, diversifying and selecting species capable of tolerating/using the oxygen. But for some other microorganisms, the oxygen levels were toxic and remained isolated in anaerobic environments. After this important oxygenation event, life has diversified even more in all three domains of life. The classification of life as we know it today comes from the work of Carl Woese, who proposed in the 90's the *three* domains of life known today: the *Bacteria*, the *Archaea* and the *Eukarya* (*15*, *16*). Based on genetic information encoded in the ribosomal DNA sequences, of all three domains, he postulated that "*Molecular sequences can reveal evolutionary relationships in a way and to an extent that classical phenotypic criteria, and even molecular functions, cannot*".

The later endosymbiotic origin of *Eukarya* and the earlier apparition of the other two domains of life, and subsequent branches, are important evolutionary concepts to understand and be able to grasp the big picture of this dissertation. I will try to explain very synthetically the relevance they have for the main subject of this dissertation in the following sections (1.2 and 1.3).

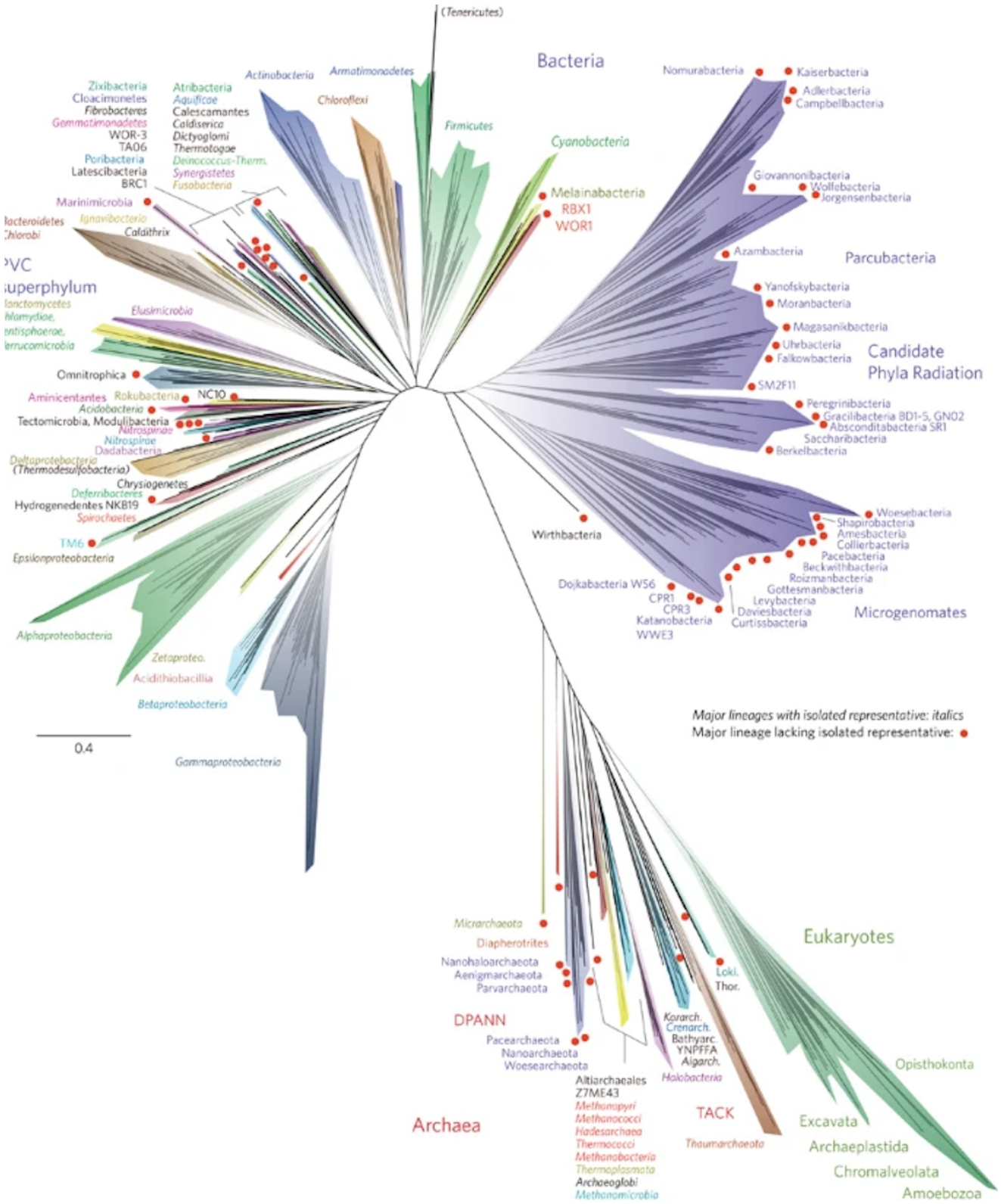


Figure 2. **The tree of life.** Current view of the tree of life based on sequenced genomes. From Hug et al. (*12*).

1.2. Mammalian holobionts

We distinguish eukaryotes, from the Greek "true nucleus", from the other two domains by having a true nucleus and defined organelles such as mitochondria, or chloroplasts in plants. They have recently evolved (in evolutionary time), to a variety of multicellular organisms where specific lineages of cells specialized in doing complementary functions for the good functioning of the multicellular organism. Mammals are an example of a multicellular organism, they appeared on Earth about 200 million years ago, derived from a reptilian order, the *Therapsida*. But at that time, life was present on our planet for at least 3.4 billion years. Today's mammals - or present in the past - represent very few branches of the tree of life and biodiversity, and have a short evolutionary time. And because they appeared much later in time, they have been always surrounded and in close contact to the microbial world. For millions of years, this microbial world has driven their evolution and speciation, setting up symbiotic relationships between/within domains (*9*, *17*). Mammals were subject to the microbial influence since they appeared on Earth, being in close contact with *Bacteria* and *Archaea*, up to the point of having life sustaining symbiotic relationships with their surrounding and inner microbial world.

Adolf Meyer-Abich, a theoretical biologist introduced the "holobiont" concept in 1943, nearly 50 years before Lynn Margulis rescued the term in 1991 (*18*) in her *endosymbiotic* theory of evolution. The holobiont word came from the Greek (hólos, "whole"), and the word biont for a unit of life, referring to the **single evolutionary** and **biological unit** that forms a **host** and its **microbiota**. Margulis proposed that mitochondria were once free living bacteria that become endosymbionts of an host cell (*10*), refining and enhancing the concept proposed in 1905 by the Russian botanist Konstantin Mereschkowski (*19*). The proposed model of endosymbiosis has been questioned and re-formulated many times because it is unclear how the protobacteria endosymbiont (*20*), become the ATP producing mitochondria in the initial archaeal host *Asgard* (*21*), proposed to be the host-to-become eukaryotic cell.

It is with these concepts of "endosymbiosis" and "holobiont" where the results of this PhD thesis can be understood under an evolutionary umbrella (Figure 3). These intimate relationships have shaped many biological systems such as the ways mammals regulate metabolism, mature their immune systems and neurological development, digest foods or activate and deactivate genes in host cells. Thanks to the advancement of technologies like mass spectrometry, DNA and RNA sequencing and other molecular tools, we start to be able to answer questions about how this multi-domain organism functions as a single evolutionary and biological unit. Where the different molecular layers of the host and its microbiota interact and maintain health or cause disease.

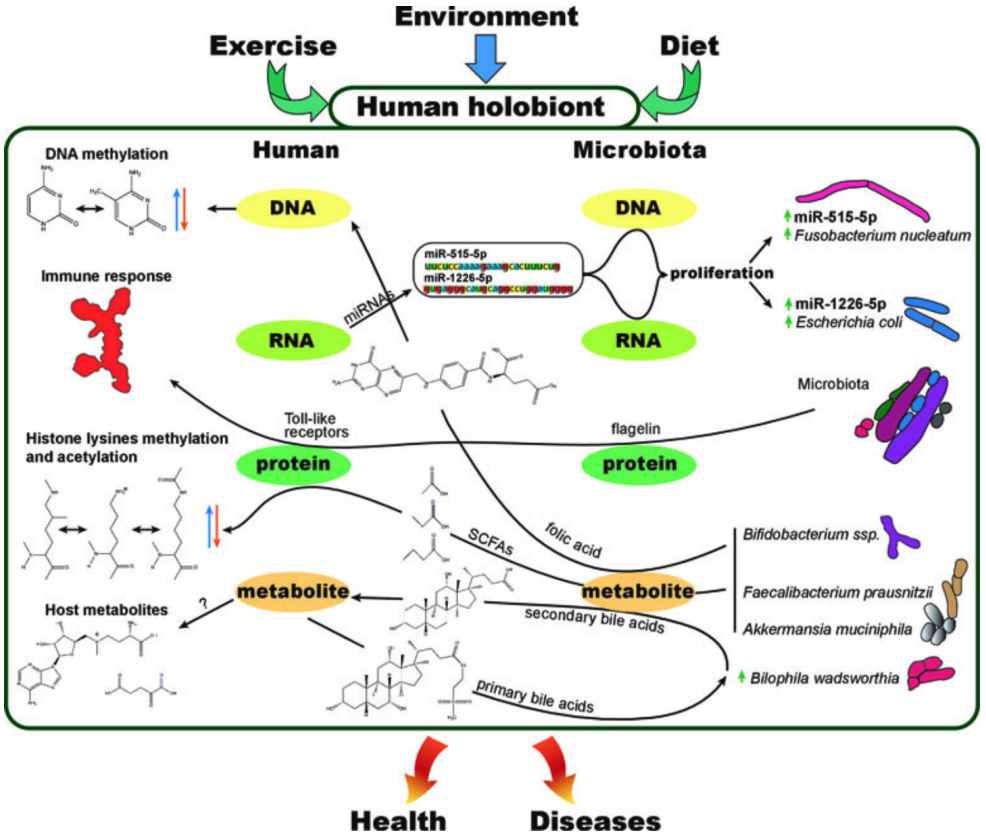


Figure 3. The human holobiont concept. Human holobiont are composed by the host, in this case the human, and its microbiota (i.e. in the gut, skin, mouth…), where all the different molecular layers from the host and the microbiota possibly interact. From Miro-Blanch and Yanes (*22*).

1.3. The "microbiota-nutrient metabolism-host epigenetic axis"

We proposed a new molecular multi-level mechanism of communication between the host and its microbiota, the "microbiota-nutrient metabolism-host epigenetic axis" (Chapter 1). This axis of communications gives one possible explanation of how the microbiota might influence host metabolism, epigenetic regulation and host gene expression. We propose in this axis that, depending on the diet and gut microbiota composition, there are epigenetically relevant metabolites ingested or produced by the host, by the microbiota or by both at the same time, that are used in epigenetic reactions in host cells to add or remove epigenetic marks and control gene expression. This axis links the diet, the microbiota metabolism, the host metabolism, the host epigenetics and its effects on gene regulation together. Metabolism from both the host and its microbiota, communicate with epigenetic marks found in the host genome, transcriptome or histones, in response to environmental stimuli to regulate gene expression. The "chemical conversation" happens when, directly or indirectly, metabolites produced by the host, microbiota or both at the same time, function as substrates, products, activators or inhibitors of epigenetic enzymes such as methyl- and acetyltransferases. Where endogenous metabolites such as acetyl coenzyme A (ac-CoA), S-adenosylmethionine (SAM) or alpha ketoglutaric acid (aKG), function as acetyl- or methyl- group donors, or as histone demethylases substrate, respectively (*23*). And where the microbiota is a key player for host homeostasis and greatly influences its metabolism (*24*) and produces exogenous epigenetically relevant metabolites such as

short chain fatty acids (SCFAs) (*25*), vitamine B12 and folates (*26*), or microbiota-derived succinate (*27*) among many other metabolites. The microbiota also helps to digest dietary fiber (*28*), produce and consume vitamins (*29*), produce and consume hormones (*30*) and neurotransmitters (*31*), help develop the host immune (*32*) and nervous system (*33*), or produces secondary bile acids (*34*). But it can also influence more systemic functions like behavior, hunger or insulin regulation (*35–37*). Microbiota composition and functionality changes dynamically in response to environmental stimuli such as diet (*38*), exercise (*39*) and lifestyle in general. And similarly, microbiota, metabolism, epigenetic marks or transcription are dynamically adjusted to external stimuli such as diet, circadian rhythms or seasonal rhythms (*40*). There are many examples of host and microbiota produced metabolites that are epigenetically relevant for the host, but we discuss some of them in more detail in Chapter 1 and 2.
In short, epigenetic marks depend on nutrients ingested from the diet, and metabolites produced by the host and the gut microbiota to activate or inhibit enzymes that dynamically add or remove chemical marks on DNA, RNA and histones, ultimately regulating transcription. I developed my PhD research for the purpose of understanding how the metabolism of the microbiota and its host, control the host epigenome and gene transcription (Figure 4).

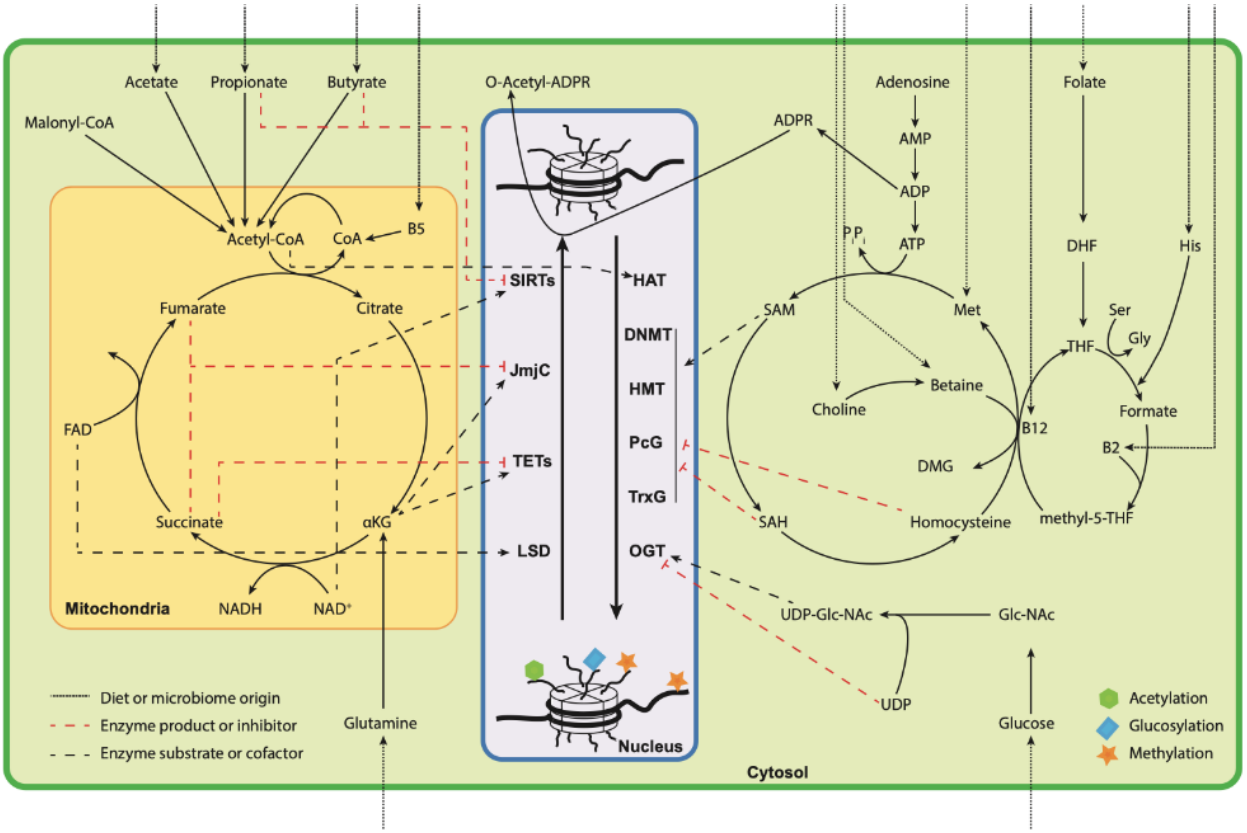


Figure 4. The nutrien-microbiota metabolism-host epigenetic axis. Main metabolites externally absorbed, produced by microbes or generated by the host metabolism that have a role controlling and fine tuning the epigenetic marks deposited in the genome, proteome or transcriptome. From Miro-Blanch et al. (Submitted, Chapter 2).

2. Metabolism

Metabolism is essential to sustain any living cell and organism. It provides the necessary energy and building blocks to create, grow and recycle cell components. Metabolism could be defined as the collection of life-sustaining chemical reactions happening in a cell or organism (*41*). Metabolism is essential for all aspects of biology and life, from fixation of CO_2 in plants to regulation of insulin in humans. The metabolism has evolved to support all functions of a cell using chemical transformations that produce energy and the necessary building blocks for the cell. The atomic composition, structure and phisico-chemical properties of metabolites vary as much as their functions in the cell. In the context of the holobiont, the host metabolism is not solely depending on the host encoded enzymes to perform metabolic reactions, microbiota enzymes and metabolism enhances the spectrum of metabolites and metabolic reactions that the host and the whole holobiont can benefit from. We can distinguish three functional types of metabolism: anabolism, catabolism, and energy transformations. These three types of metabolism apply to sugars, lipids, amino acids, nucleotides, vitamines, proteins and macromolecules such as DNA, chromatin or protein complexes among others. Ultimately, these three types of metabolic reactions are responsible for maintaining cell and organism homeostasis.
In this section (2) I describe what metabolism is, explaining the roles of metabolism in human disease and its connection to epigenetics and microbiota.

2.1. Cell metabolism and homeostasis

Metabolism can be described as a web of interconnected metabolites and molecular reactions, structured in metabolic pathways, where the product of one reaction is connected to the following reaction as a substrate. These metabolic pathways are heavily regulated by different molecular mechanisms at different molecular levels. From epigenetic mechanisms regulating metabolic pathways, to DNA methylation affecting gene transcription, or with non-epigenetic mechanisms such as the phosphorylation of the protein that catalyzes the metabolic reaction, or the abundance of a substrate that limits the metabolic reaction (*42*). For example, metabolic pathways are often encoded by many genes that can be coordinately expressed or repressed and control their metabolic pathway (*43*). The regulation of a specific pathway is determined by the cell or organism status, and what the cell needs, in terms of substrates, enzymes and conditions, to accomplish its functions and stay in homeostasis (Figure 5) (*44*).

From the Greek "homoios", meaning "similar", and "stasis" meaning "standing still", appears the concept of "staying the same" to which homeostasis refers. The homeostatic status of a cell is when physical and chemical conditions are maintained constant with a small range of variation for optimal cell functioning (*45*). Cells try to maintain a balanced metabolism to not suffer from stress, and maintain constant concentrations of metabolites with delimited fluctuations. But, to maintain cell homeostasis, many metabolic reactions are happening at the same time, sometimes compartmentalized in different cell organelles or tissues, or separated in time. As previously mentioned, there are three types of metabolic reactions that maintain cell homeostasis: catabolic reactions, anabolic reactions and energy transformation reactions. Catabolism breaks down bigger molecules to release energy and

generate precursors for other metabolic reactions. Anabolism consumes energy and combines smaller molecules to generate larger biologically functional molecules. An energy transformation is when for example an eukaryotic cell uses a gradient of protons to create energy in the form of ATP in the respiration chain through oxidative phosphorylation in the mitochondria (*46*). These reactions are modulated by enzymes that catalyze the chemical reaction between two or more metabolites.

Healthy living cells control metabolism through a plethora of molecular mechanisms such as but not limited to: differential gene expression, histone and DNA modifications, chromatin structure, specificity of transcription factors, alternative splicing, protein folding, abundance of substrates, products and enzymes, protein isoforms, phosphorylation of proteins…obtaining the right metabolite at the right place, in the right abundance and at right time. In general, the final effector which controls the molecular mechanisms producing a metabolite, ends up being an enzyme, which catalyzes the reaction, lowering the energy needed and accelerating the reaction. Enzymes are usually proteins with different amino acid sequences that have an evolutionary conserved amino acid sequence at their catalytic center. The composition of the sequence of amino acids forming an enzyme defines their folded three-dimensional structure and catalytic activity. Another group of macromolecules with catalytic activities are ribozymes, RNA structures with catalytic activity that regulate alternative splicing, translation and other RNA related biochemical reactions. Ribozymes have dual functions being a catalyst and carrying genetic information at the same time (*47*).

Enzymes are regulated by different cell mechanisms to control their activity. First, the enzymatic activity can be regulated by the availability of the elements in the reaction, mainly substrates and products which often limit the enzymatic reaction. Then the enzyme itself can be regulated at different molecular levels:

1. Gene expression: enzyme abundance can be regulated by controlling the rate at which the gene coding for the enzyme is expressed, which is controlled by epigenetic mechanisms, turning on and off the gene.
2. Transcript translation: once we have the transcript of a particular enzyme, the translation to the protein sequence can be regulated by the type of splicing and alternative splicing, which can give different isoforms of the enzyme with slightly different enzymatic activities.
3. Post translational modification: when we have the spliced and active isoform of the enzyme, this protein can be modified at virtually any residuous of their amino acid sequence to change their folding conformation and enzymatic activity, for example adding a phosphorus group.
4. Localization: in a eukaryotic cell, metabolic reactions are compartmentalized and separated from other reactions to be able to increase the number of simultaneous reactions, and control when and where these reactions happen.
5. Inhibition: there are competing molecules that can regulate the catalytic activity by creating strong or weak bonds to the enzyme, slowering, interfering or stopping the enzymatic metabolic reaction.

In brief, cell metabolism is regulated by different epigenetic and non-epigenetic mechanisms that ultimately aim to maintain homeostasis. The homeostatic state might change depending on the cell cycle, tissue or organism needs, and redirect and regulate different pathways to allow different metabolic reactions to happen at the same or different times, in the same or different locations to accomplish specific functions. Where metabolism produces the building block and energy to sustain

any living cell or organism, and has evolved molecular mechanisms at different levels to regulate its reactions depending on the cell or organism fate.

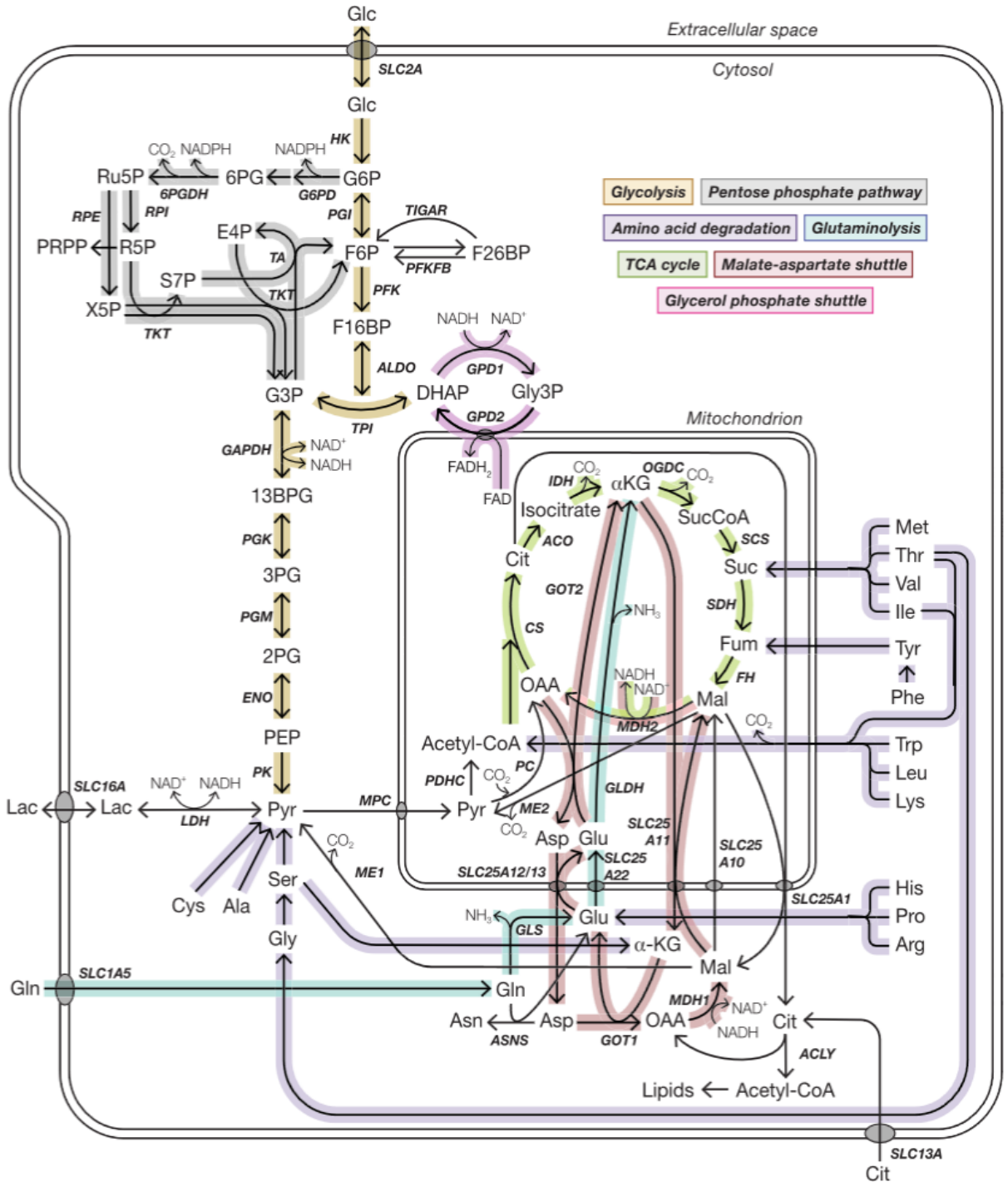


Figure 5. Central metabolism homeostasis simplified. Principal metabolite pathways involved in cell homeostasis. From O'Brien et al. (*44*)

2.2. Metabolism, diseases and low grade inflammation

Metabolism is at the root of many diseases, not only in metabolic diseases like type I and II diabetes, metabolic syndrome or obesity, but also in cancer, autoimmune or neurodegenerative diseases among many others. Keeping a balanced metabolism is necessary to maintain health, and when there is a disruption in the metabolism, a disease phenotype or a disease signature might appear (*48*). For a period of time many common diseases were considered only in terms of mutations, gene regulation or cell differentiation, losing the interest in small molecules. But, in the past decades there has been an increasing curiosity in the role of metabolism and metabolites considering gene regulation or cellular differentiation from a metabolic perspective (*49–52*). This has allowed us to gain more mechanistic insight to disease, for example, performing genetic knockdown or cell differentiation experiments and measuring metabolites to support the genetic results with a metabolic and genetic explanation of the phenotype (*53*).

Metabolic perturbations in diseases can have many origins and consequences, for instance they can originate from a somatic mutation in genes such as KRAS, TP53 and SMAD4 affecting immune-related signaling pathways in the pancreas, ultimately leading to pancreatic cancer patients suffering from diabetes (*54*). Another possibility of many, is that a post translational modification such as histone lysine acetylation regulates the activity of metabolic enzymes involved in glycolysis, gluconeogenesis, tricarboxylic acid cycle (TCA) or fatty acids metabolism (*55*), being disrupted in many diseases. These two examples highlight the interconnection of genetic and epigenetic events with metabolic consequences, but the opposite can also happen. Where a metabolite such as methionine is too abundant, increasing the metabolic reaction that forms S-adenosylmethionine, the main methyl donor in methylation reactions, favoring the transfer of methyl groups to the DNA and silencing gene transcription (*56*). These two similar scenarios where a gene mutation or an epigenetic mark could have metabolic consequences, or the opposite, where some metabolic cues could affect the expression of a gene through regulating an epigenetic mark, might contribute to metabolic disorders such as diabetes, obesity, cancer and other diseases, and is one of the main focuses of this PhD research.

Obesity is one of the leading causes in the onset of many diseases, and has been linked to a disrupted metabolism in many diseases such as type I and II diabetes, inflammatory bowel diseases, metabolic syndrome or cancer, among others (*57*). Often obesity is associated with a low-grade chronic inflammation, specially in adipose tissue where more immune cells infiltration and releasing of signaling interleukins, leads to a low-grade chronic inflammation (*58*). Another cause or consequence of obesity, might be the dysfunction of mitochondria, where the accumulation of reactive oxygen species (ROS) can cause oxidative stress, due to excessive nutrients supply in obesity (*59*), leaving a permanent molecular signature in adipose tissue not recovered after a life-slytle intervention (*60*), disrupting the adipose endocrine function (*61*), and many other hallmarks of a diseases status are associated to mitochondrial dysfunction and obesity (*62–65*). But recently, the low-grade chronic inflammation has been proposed to be a consequence of a disrupted gut microbiota in the obese and non-obese patients suffering from metabolic diseases (*66–68*).

In recent years, the relationship between some metabolic diseases and gut microbiota has been the object of many research projects trying to understand the diseases from a microbiota point of view.

Many researchers have found in a variety of diseases, a distinct microbiota signature between the individuals suffering the diseases, from healthy control subjects. But often, results are inconsistent due to the huge microbiota variability between individuals. For example: obesity has been linked to a disruption and lower gut microbiota diversity in obese states compared to lean ones (*69*); type I diabetes has been found to be dependent on the interaction of gut microbiota and the immune system (*70*); the interaction of gut microbiota and host genotype can contribute to the development of metabolic syndrome (*71*); a link between colorectal cancer and gut microbiota has been also made (*72*): but other type of cancers and even the efficacy of cancer drugs seem to depend on the gut microbiota (*73*); cardiometabolic diseases such as ischemic heart disease are also dependent on the microbiota (*74*); inflammatory bowel diseases such as Crohn's disease or ulcerative colitis have different microbiota and epigenetic signatures in the colon (*75*); and the microbial activity has been associated to signaling and disruption of mitochondrial functions in the host in colorectal cancer (*76*); but the opposite might also be happening, where mitochondria drives the gut microbiota composition that, in turn, can influence disease outcome (*77*); anxiety and depression have been also been linked to microbiota composition, where the supplementation of a probiotic augmented the expression of GABA receptors involved in these mental disorder (*78*).
In summary, disruption of metabolism is linked to many metabolic diseases such as obesity, diabetes or metabolic syndrome, and non- metabolic diseases such as cancer, neurodegenerative or autoimmune diseases. However, the co-regulation of host and microbiota metabolism has been involved in the onset of many diseases, having all a metabolic component and linking microbiota and metabolism in diseases.

2.3. Host-microbiota metabolic interactions

In this thesis I am investigating two major events related to metabolism: how metabolism impacts the epigenome of the host; and how the interaction between the host and microbiota metabolism occurs.

2.3.1. **Metabolism and epigenetic interactions**

The interaction between the metabolism and epigenetic marks can occur when epigenetic marks such as DNA methylation or histone acetylation, are regulated by the availability of S-adenosylmethionine or acetyl-Coenzyme A to transfer a methyl or acetyl group to the DNA or histone. The high number of possible residues that can be modified in the chromatin (in the DNA and histone tails), implies appropriate levels and regulation of metabolites in the cell, to ensure availability of substrates to supply to the modifying enzyme when needed. The consumption of key metabolites by DNA or histone-modifying enzymes is consistent with the concept that these enzymes are sensing the metabolic state of a given cell, and can respond to it (*79*). However, not only the global level of the substrate metabolite is the sole determinant of its enzymatic activity, the rate, spatial and time distribution of the metabolite might be limiting factors that regulate the reaction (Figure 6) (*23*). Metabolite availability and cell fate will determine the kind of epigenetic mechanism needed at a particular cellular phase and location, to ensure appropriate regulation of transcription and translation.

In recent years the number of known epigenetic modifications has increased considerably, thanks to advances in technology like next generation sequencing (NGS) and mass spectrometry (MS). This has

helped to discover new modifications on macromolecules like DNA or histones, different from the three previously known canonical modifications, DNA and histone methylation, and histone acetylation. These new epigenetic modifications implicate other metabolites such as lactate, succinate or butyrate, coming from different metabolic origins (*80–82*), together with the rediscovery of the RNA as a target macromolecule that can be modified too. This has been demonstrated to have more than hundred different modifications and a key role in regulating gene expression (*83*). Metabolites connect different parts of the metabolism with epigenetics marks, allowing to have a more diversified system of molecular mechanisms regulating gene expression and translation. Where metabolites exert different functions in epigenetic reactions, depending on their abundance, time and localization in the cell, being substrates, activators, inhibitors or products of the epigenetic reaction.

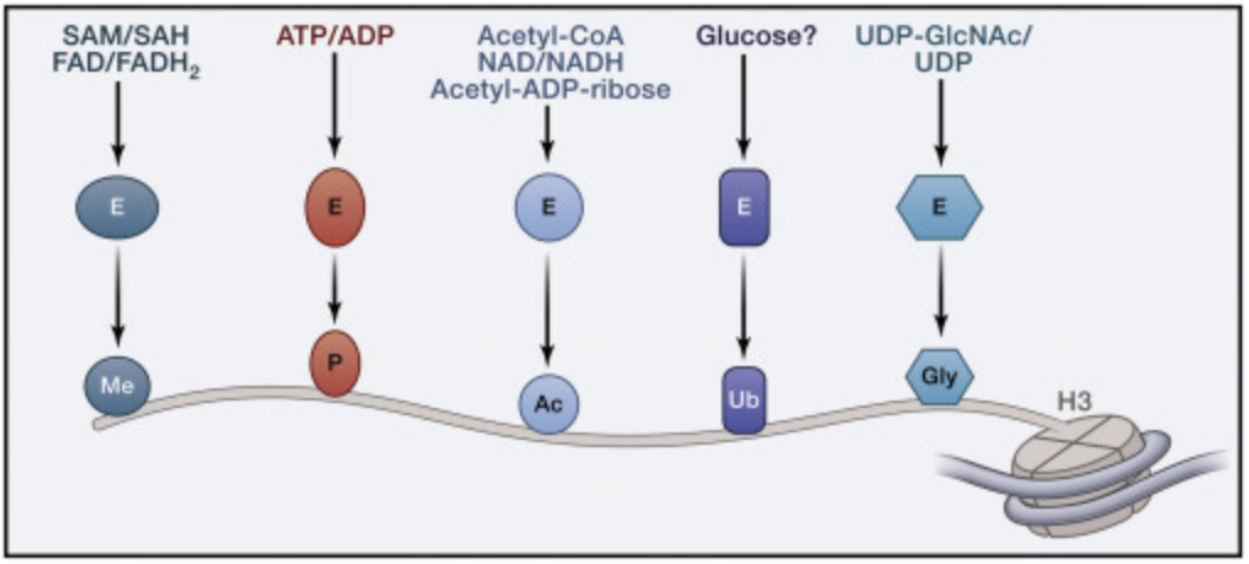


Figure 6. Metabolism-epigenetic axis. Chromatin remodeling enzymes "sense" cellular metabolism and react to environmental and metabolism stimuli. From Katada et al (*23*).

2.3.2. **Host and microbiota metabolism interactions**

Microbiota play an important role in the host metabolism. The microbiota is the community of microorganisms (bacteria, archaea, fungi, viruses and microscopic eukaryotes) that live in a specific location or ecosystem (see section 3). Mammals host a rich community of microorganisms in many places of their bodies where they have important biological functions. The largest community of microbes in a mammalian host is located in the colon and intestinal tract. These microbes contribute to the metabolism of the host by producing a myriad of small molecules that are distributed through the circulating system (*84*). They contribute in training the immune system (*85*), mature the nervous system (*86*) or produce some host peptide hormones among other functions (*30*). In a metabolic context, these microbes can produce and consume hormones, neurotransmitters, secondary metabolism products, TCA intermediates, vitamines, short chain fatty acids (SCFAs)…. Metabolites produced by the gut microbiota metabolism are usually beneficial for the host and fill the metabolic niches left empty by the host encoded enzymes.

SCFAs are probably one of the most studied metabolites produced by the gut microbiota, and at the same time one of special relevance for this PhD thesis. Acetate, butyrate and propionate are known to be implicated in many host functions: they are the primary source for colonocytes (*87*), have immunomodulatory functions (*85*), are histone deacetylases inhibitors (*88*), function as acetyl donors (*89*), participate in brain signaling (*90*), repress PPAR-γ expression (*91*), among many other functions. SCFAs make the link between microbiota, metabolism and epigenetic marks because for example, butyric acid and acetic acid are potent histone lysine deacetylase inhibitors (*88*, *92*, *93*), and propionic and acetic acid are used to directly modify histones (*94*) and regulate gene transcription.

Another metabolic function of the host microbiota is to perform a secondary metabolism on host primary bile acids (*34*), they transform primary host bile acids to new bile-acids conjugates. Those new conjugates have a signaling and immunomodulating function (*95*). Microbes also produce and consume significant amounts of intermediates of the TCA cycle such as succinate (*27*), fumarate or alpha-ketoglutarate, and these metabolites play part in the homeostasis and signaling pathways of central metabolism (*96*).

The host and microbiota compete for vitamins and essential amino acids (*97*), but often, the microbiota is producing vitamins in benefit to the host. Microbiota produces vitamin K and some B vitamins such as cobalamin, biotin, folates, riboflavin, pantothenic acid or thiamine (*29*, *98*), essential cofactors for many reactions. Microbiota might also contribute to the metabolism and flux of metal ions such as iron, manganese or magnesium, among others. It controls the iron levels and modulates liver fat accumulation contributing to non-alcoholic fatty liver disease (NAFLD) (*99*).

Host sexual hormones are also under the influence of microbiota metabolism, where specific taxons provide the necessary signals that control a sex-biased predisposition to diseases (*100*). For example, the levels of testosterone in the host are determined by the type and composition of the gut microbiota (*101*, *102*).

Neurotransmitters are linked to the composition and diversity of the gut microbiota, being microbes capable of altering mood and behavior by releasing or consuming specific neurotransmitters (*35*). Where the vagus nerve provides a highway for metabolites (neurotransmitters included) to go from the enteric nervous system to the brain, what is known as the gut-brain-axis (*103*). And where specific species of bacteria such as *Coprococcus spp* are found to be depleted in depressed patients (*104*).

Together with the metabolic reactions in the host, summed to the abundance of metabolites released by the gut microbiota, and the enzymatic epigenetic readers, writers and erasers and other enzymes, we have a complex metabolic and molecular ecosystem that regulates gene expression and homeostasis in the host (Figure 3-4). And concepts in the fields of molecular ecology or biodiversity might provide us with a better framework where we can understand the link between nutrients, metabolism, microbiota and epigenetics in the context of a holobiont (organism with multiple bionts).

3. Microbiome

The microbiota is composed by bacteria, archaea, fungi, viruses and by microscopic protists (*105*). But the higher abundance, genetic plasticity and genetic diversity of bacteria makes them the most important group of microbes for a mammalian host. Most microbes reside inside the intestinal tract, where the colon possesses the highest abundance and diversity of microbes. They help the host to digest dietary fiber, train their immune system, mature the nervous system, produce and degrade hormones, neurotransmitters, vitamines, influence host behavior….and new functions are discovered almost every week. Microbes are present in most organs and tissues in the mammalian host, from lung (*106*), possibly in the brain (*107*), or in the placenta (*108*). In each location they may play essential roles to maintain good functioning of the tissue and cells, and prevent colonization by pathogens. Mammals acquire the initial microbiome at birth - or even before birth during pregnancy - and this process is very dynamic and depends on many factors such as mode of birth, feeding, environment or diet (*109*). Microbiota influences our health in many ways, and it has been associated with many pathologies and has numerous health benefits.
In this section (3) I introduce and review key concepts about the microbiota, I describe how we classify microbes, use them in food, their ecological importance and how we acquire the microbiota and how it changes through aging, and discuss the role of the microbiota in health and disease, and strategies to modulate it.

3.1. Microbiota and microbiome

Due to their higher abundance in the intestinal tract and their enormous genetic functional potential and plasticity, scientists refer most of the time to bacteria and archaea when they talk about "microbiome" or "microbiota". Strictly speaking, the microbiota is composed of many other microorganisms such as viruses, fungi and other microscopic organisms and not solely of bacteria or archaea. But, I am going to use the terms "microbiota" or "microbiome" interchangeably, to refer mainly to the *Bacteria* and *Archaea* domains of life, unless explicitly stated. And the same applies to the manuscripts presented in section 9-11 and the whole PhD thesis.

3.1.1. **Microbiota and microbiome**

The terms "microbiome" and "microbiota" are used interchangeably in many scientific articles referring sometimes to the microorganism present in a defined ecosystem and associated traits. However the term "microbiota" is defined as the assembly of microorganisms belonging to different domains and present in a defined ecosystem. And the term "microbiome" is defined as a characteristic microbial community present in a macro-ecosystem and their associated genes. Involving the formation of specific niches, a microbial community integrated in a dynamic and interactive manner across time and space into other macro-ecosystems such as the mammalian gut (*110*).

3.1.2. **Microbial taxonomy**

The classification of *Bacteria* and *Archaea* is not an easy endeavor and has changed over time. Antonie Van Leeuwenhoek first discovered bacteria in the year 1676, and called them "animalcules", meaning small animals. He is considered the "father of microbiology" and discovered bacteria by looking at scrapings of his own teeths, and ones from others. He extensively described his observations in letters to the Royal Society in the UK, observations made with his own constructed lenses, the first rudimentaries microscopes. But, microbial taxonomic classification did not start until the 19th century and still continues to evolve today. The first methods of classification were based on morphology, growth rates and growth requirements and pathological potential (*111*). At the beginning of the 20th century, the classification was based on biochemical assays and data available about physiology and morphology of microorganisms. Later in the 1960-80's, with the discovery of the DNA molecule, the classification was based on DNA-DNA hybridization, as well as based on their chemical similarities and dissimilarities of the compounds produced by the microorganisms. But, since the invention of the PCR in 1983 and the subsequent development of Sanger sequencing, the classification of *Bacteria* and *Archaea* has been based on the genotypic analysis, multi-locus analysis and sequencing analysis of few hypervariable genes or the whole genome (*112*). Despite the use of next generation sequencing technologies nowadays, it is still challenging to standardize a classification method because of the high variability in genetic information, and high rates of horizontal gene transfer between species (*113*) within the domains of *Bacteria* or *Archaea*, that make the assignment to a particular taxon very difficult, and sometimes temporary (*114*).

The most reliable method of *Bacteria* and *Archaea* classification is sequencing the whole genome followed by the sequencing of the 16S gene in one or few of the hypervariable regions. The advantages of whole genome sequencing (WGS) relies on the increased information obtained from sequencing most of the genome present in a sample: identification of significantly more bacterial species, more accurate assignment of species, increased richness and diversity detection, and with the added value of being able to detect, viruses, fungi and protozoa, as well to allow direct functional analysis. But the 16S technique also has some advantages compared to the whole genome strategy, sequencing only a single region of the genome: it is cost effective, data analysis is performed with established pipelines, and there is a large amount of data to use as reference (*115*). With these two techniques (WGS and 16S), scientists analyze biological samples to study the composition and functionality of the microbiomes present in samples like a swap of human skin, a sample of mouse feces, a soil sample or a deep-sea water sample collected underwater. These techniques help to better classify microbes and might help to understand and observe the effect of a perturbation, a treatment, or simply observe and characterize the microbial features of a sample.

3.1.3. **Microbial fermentation**

Fermentation is the primary source of energy for microbes, in the form of ATP in the absence of oxygen, and it is one of the oldest pathways in metabolism since life appeared on Earth. Fermentation can be carried out by bacteria, archaea and eukarya and is present in very different environments, from freshwater sediments to the rumens of cattle, or in the human gut (*89*, *116*, *117*). Fermentation has been used by humans to preserve and transform foods since the Neolithic. It relies on the biological

activity of microorganisms where enzymatic chemical changes in organic molecules occur, mostly performed by bacteria, archaea and fungi. Lactic fermentation and alcoholic fermentation are the two most used types of fermentations in human history, to preserve, transform and produce different types of food (*118*). Wine, beer or bread are produced by the fungi *Saccharomyces cerevisiae*, and lactic acid producing bacterias (LAB). Cheese, yogurt and kefirs are produced by the genus *Lactococcus* and *Lactobacillus*, as well as the fermentation of some vegetables (*118*). These microbes are probiotics, living microorganisms that when ingested enrich and increase the gut microbiota diversity of the host. I enter more in depth on the uses of probiotics and prebiotics in section (3.2.).
But fermentations also occur inside the mammalian intestinal tract, where restricted concentration of oxygen forces the microbiota to fermentate nutrients to obtain energy. Gut microbiota produce short chain fatty acids (SCFAs) through the fermentation of undigested polysaccharides ingested by the host, that cannot be degraded by the host because the lack of encode appropriate enzymes to digest these polysaccharides (*90*, *119*, *120*).

3.1.4. **Microbial ecology**

The ecological rules that shape the microbial diversity in the mammalian gut apply to mutualism and pathogen relationships (*121*). These relationships are driven by ecological forces that shape the microbial community in its continuous evolution and ecological succession. An ecological succession is the process by which an ecological community, for example the gut microbiota, changes its structure of species within the ecological community over time, favoring fitness. We can distinguish two types of ecological succession: a primary ecological succession and a secondary ecological succession. A primary ecological succession happens when there is a colonization of an area that has not been occupied previously, such as the guts of a germ-free mice (*122*, *123*), in a mammalian host context we could think about a newborn, where birth is supposed to be the first contact with microbes and those colonize all possible locations of the newborn. A secondary ecological succession happens when there is a colonization of an area that had previously been occupied, but due to a disturbance or removal of the previous community, a new ecological succession starts(*124*). In a mammalian host context, we could think about the disturbance that exert to the existent community in the gut, the treatment with antibiotics (*125*). Secondary successions are strongly influenced by the composition of the existing community previous to the disturbance, where residual characteristics from the initial community make the secondary succession more rapid and reach the climax community faster (*126*).

A typical ecological succession in a place where there is space and resources for microbes to grow goes similar to the following: initially, there are several niches available to be inhabited by microbes, and first pioneer microbes arrive and ocupate the different niches, then the microbial succession starts until an ecological mature community or climax community is reached, where fitness determines which species stays and which ones do not. The climax community is adapted to environmental conditions and might stay stable if no major disruption happens, it is the permanent or final stage of an ecological succession (*127*, *128*).
More recently, ecologists have demonstrated in microbial communities an interesting concept where, under fixed experimental conditions, high biodiversity and fluctuations reinforced each other. And where simple community-level features dictate emergent behaviors of the ecological community. The ecosystem transitions between three distinct stages, from stable equilibrium in which species coexist to

partial coexistence or to emergence of persistent fluctuations in species abundances (*129*).
The relationships that drive this ecological succession are driven by interactions between pairs of species (microbe-microbe, or host-microbe), and are one of the following relationships in Table 1:

Relationship	Species 1	Species 2
Commensalism	+	0
Mutualism or cooperation	+	+
Neutralism	0	0
Competition	-	-
Predation	+	-
Amnesalism	-	0

Table 1. Symbiotic relationship in an ecosystem. Extracted and adapted from the Nature Knowledge Project.

Three robust clusters of bacterial species that are not continent or country-specific have been identified as enterotypes (*130*). Each enterotype has specific microbial compositions and functionality, and its variation is generally stratified rather than continuous (Figure 7.c). Each strata having specific ecological functions and composition, well differentiated from the contiguous strata, and possibly used as diagnostic or prognostic biomarkers. The authors also remark that having different enterotypes might not be a unique feature of human microbiomes, but also present in other animal species. Another recent study has shown the existence of mixed-membership enterotypes, using a different approach the authors have proven how a nested structure is present, implying both the microbes and the hosts being one of the two: generalists or specialists (*131*). And shown how a healthy gut microbiota is a combination of both, generalists and specialists robustly stratified (Figure 7.a-b).

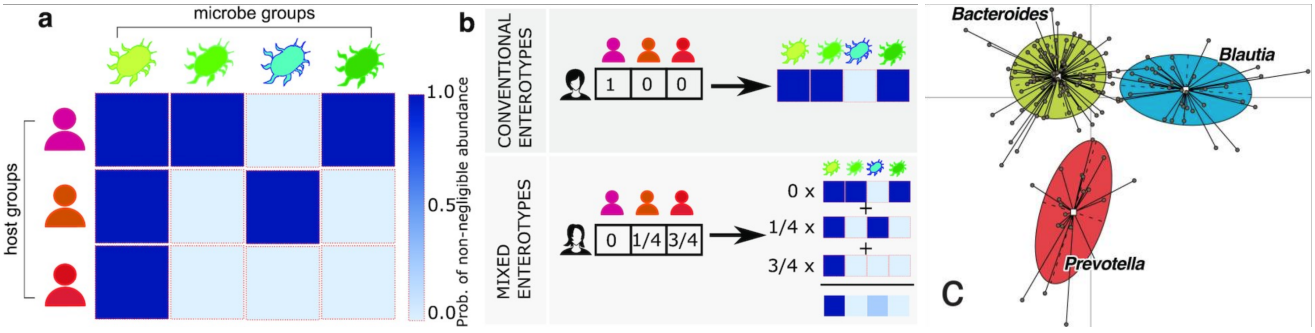


Figure 7. Mixed-enterotype model and classical enterotype model. The mixed enterotype model proposes nestedness as a source of stability with generalists and specialist species within the enterotypes, while the classical enterotype model proposes three defined enterotypes of taxonomic composition in the microbiota. From Cobo-López et al (a,b)(*131*). and from Arumugam et al (c)(*130*).

To measure the abundance and count the number of microbial species within a sample, or between samples different biodiversity indices such as alpha and beta diversity, evenness or shannon index among others, have been created. The alpha diversity and beta diversity were first introduced by Whittaker (*132*). The total species diversity in an ecosystem is called the gamma diversity is influenced by two other diversity indices: the alpha and beta diversity. Alpha diversity refers to the average of species present in a specific individual or sample, and beta diversity refers to the diversity differences between two or more individuals or samples. Where low beta diversity is usually observed in samples with very similar microbial composition. Chao 1 or Jackknife indexes measure the richness present in a sample (total number of species) in slightly different ways, to estimate the alpha diversity of a sample. Another measure that can be used is the evenness indices, that measure how evenly the relative abundances are distributed, and assess how stable and resilient an ecosystem is (*133*). Most used alpha diversity indices can be observed in Table 2.

Name	Alpha diversity index	Relation with ${}^{q}D$	Class
Observed richness	S^{obs}	${}^{0}D$	Richness
Chao1	$S^{Chao1} = S^{obs} + \frac{f_1 (f_1 - 1)}{2(f_2 + 1)}$		Richness
1st-order Jackknife	$S^{Jackk1} = S^{obs} + f_1$		Richness
2nd-order Jackknife	$S^{Jackk2} = S^{obs} + 2f_1 - f_2$		Richness
ACE	$S^{ACE} = S^{abund} + \frac{S^{rare}}{C^{ACE}} + \frac{f_1}{C^{ACE}} \cdot \gamma^{ACE 2}$		Richness
Hill numbers	${}^{q}D = \left(\sum_{i=1}^{S^{obs}} p_i{}^{q} \right)^{\frac{1}{1-q}}$		Diversity
Berger–Parker	$BP = \frac{1}{max(p_i)}$	$\frac{1}{{}^{\infty}D}$	Diversity
Rényi entropy	${}^{q}RE = \frac{1}{1-q} ln \left(\sum_{i=1}^{S^{obs}} p_i{}^{q} \right)$	$\ln({}^{q}D)$	Diversity
Inverse Simpson	$IS = \frac{1}{\sum_{i=1}^{S^{obs}} p_i{}^{2}}$	${}^{2}D$	Diversity
Gini–Simpson	$GS = 1 - \sum_{i=1}^{S^{obs}} p_i{}^{2}$	$1 - \frac{1}{{}^{2}D}$	Diversity
Shannon entropy	$H = - \sum_{i=1}^{S^{obs}} p_i \cdot ln (p_i)$	$\ln ({}^{1}D)$	Diversity
Tail	$T = \sqrt{\sum_{i=1}^{S^{obs}} (i-1)^2 \cdot p_i}$, with $p_1 \geq p_2 \geq \cdots \geq p_S$		Diversity
EF	${}^{q}EF = \frac{{}^{q}D}{{}^{0}D} = \frac{{}^{q}D}{S^{obs}}$		Evenness
RLE	${}^{q}RLE = \frac{ln({}^{q}D)}{ln({}^{0}D)} = \frac{ln({}^{q}D)}{ln(S^{obs})}$		Evenness
Pielou	$P = \frac{ln({}^{1}D)}{ln(S^{obs})}$		Evenness

Table 2. Alpha diversity indices. Principal alpha diversity indexes and the mathematical formula from which they are derived. From Finotello et al. (*134*)

3.1.5. **Acquisition and evolution of the human microbiome**

Perinatal age

Until recently, the mother womb and placenta were considered sterile in humans, being birth the first contact with the microbial world for babies. However, there is a small body of evidence on the presence of microbiota in the fetus during pregnancy, where a limited number of species are detected in fetal intestinal samples (*135*). The key question then is if there are different health outcomes for infants that have been in contact with bacteria during pregnancy, from the ones that did not have any contact, or had contact with different species (*108*, *136*, *137*)?

Birth mode greatly influences the seeding, presence and posterior ecological succession of bacteria in the newborn, changing their acquisition and structure of the initial microbiota in multiple body locations (*138*, *139*). For example, newborns delivered by c-section tend to have a gut microbiota composition more similar to their mother skin than to a vaginal microbiota composition observed in a vaginally delivered infant (*140*). Dominguez-Bello et al. have shown that we can partially "restore" the microbiota composition of birthing newborns by c-section by performing a mother vaginal swab right after the birth (*141*). The initial microbial transmission from mother to infant is a dynamic process, where the mother gut microbiota is the most important source of stable microbiota in the newborn (Figure 8.a) (*109*). The use of intrapartum antibiotics prophylaxis or in the perinatal period by the pregnant mother has been linked to reduced gut microbial diversity in their infants and increased presence of beta-lactamases resistance genes (*142*, *143*). Other studies have demonstrated that the effect of antibiotic treatment in the newborn is time and dose-dependent, being critical for the development of a healthy microbiota and having long-lasting effects (*144*).

Breastfeeding and microbiota

Another important factor for the establishment and normal development of the newborn gut microbiota is whether the infant feeds on their mother's breast milk, pasteurized human milk or if it is formula fed (*145*)(Figure 8.b). Initially, the breast milk was thought to be sterile, but we know today that a complex microbial community is present in the milk to help establish the infant gut microbiota (*146*). There are two hypotheses that might explain how microbes arrive to the milk: the entero-mammary transfer of maternal gut microbiota, or the retrograde hypothesis proposing the inoculation to the breast from the infant mouth (*147*). These two hypotheses might both be true, because there is strong evidence for both fenomena to be happening at the same time (*148*, *149*). Factors like human milk oligosaccharides (HMO), hormones, immune cells, and antibodies could be modulating the milk microenvironment constraining the ecological niche to establish infant gut microbiota (*150*). HMO were first thought to function as feeding substrates for established infant gut microbiota species, however their role goes beyond that. HMOs serve as prebiotics, nutrients for bacteria, giving certain species of bacteria of the genus *Lactobacillus* and *Bifidobacterium* competitive advantages over pathogenic species, keeping pathogens in check (*151*). They also serve as antiadhesive antimicrobials, preventing the binding of *Campylobacter* species to the 2H-antigen in epithelial cells (*152*). HMOs are also known to be potent modulators of intestinal epithelial cells responses to environmental stimuli, being immune modulators, protect against NEC in preterm infants and improve brain development, among other benefits (*153*).

Although HMOs are an important factor to establish gut microbiota in the infant, the establishment and stability of the infant gut microbiota has been demonstrated to be strongly dependent on the microbiota passed to the infant by breast feeding practices (*154*). Interestingly, infant sex-dependent breast milk microbiota composition has been reported, supporting the retrograde hypothesis (*150*). When solid food is introduced to the infant, a big change in microbiota composition happens. Depending on cultural background, age, geography and nature of the solid food, changing from exclusively breast feeding to introducing solid foods and weaning has a strong effect on the infant gut microbiota composition (*155*). Recent studies have suggested that microbiota development in the infant may take longer than previously thought (*156*), continuing evolving after the age of three and through adolescence in contrast to what we previously knew (*157*).

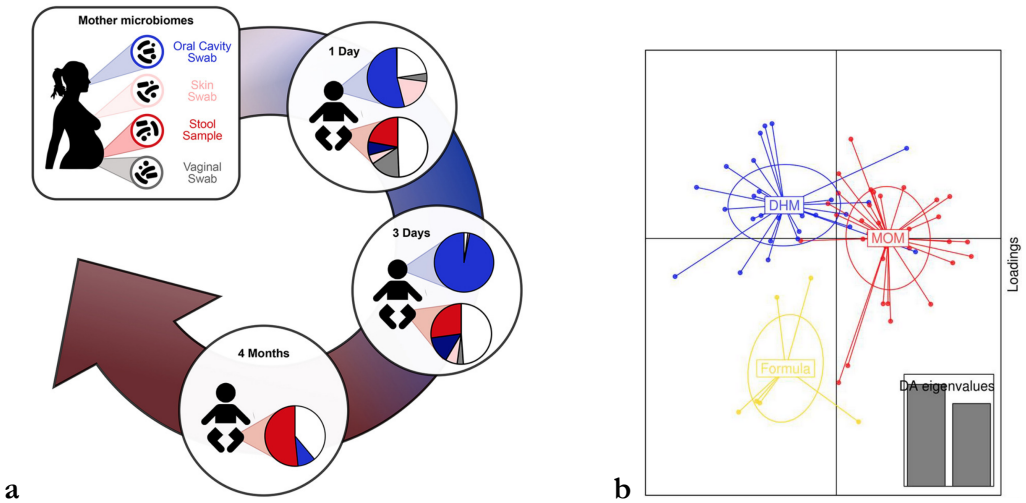


Figure 8. The perinatal microbiome. The ecological succession in early life (a). From Ferreti et al. (*109*). (b) Discriminant analysis of principal components of the microbial composition in breastfeed (red), formula fed (yellow) and pasteurized donor human milk fed (blue) babies. From Parra-Llorca et al (*145*).

Age and geography

Microbiota composition exhibits changes through the different stages and along the whole life, from childhood to adolescence, to adulthood or to senescens, a continuous ecological succession happens (Figure 9) (*139*). During the perinatal period, and until weaning and introduction of solid foods, microbiota of neonates have lower diversity and are relatively dominated by the phyla *Proteobacteria* and *Actinobacteria,* and are later dominated by *Bacteroidota (Bacteroidetes)* and *Bacillota (Firmicutes)* phyla (*158*). By one year, each newborn has a unique and distinct microbiota, and by three years, we thought that the microbiota resembles very much the one of an adult in terms of composition and diversity (*157*), however new evidence says that microbiota continues to evolve after these age and through adolescence (*156*). After this early window to establish the infant microbiome, environmental exposures during life are shaping gut microbial ecology, and determine diversity and functionality (*157*). For example, you are exposed to very different microbes if you live in a small farm in the countryside, or if you live in a big city such as London, in the UK. The type of microbes present, the number of persons or animals that you interact with or the pollutants to which you are exposed in a city are radically different from the ones you can find in a farm with animals, with much less people and more contact with nature. Or if you live in the Sahara desert or in the Amazon rainforest, the type of foods you have

access to, the surrounding nature or the clima, also have a strong impact on the microbiota. It has also been demonstrated that strong differences exist between underdeveloped countries and developed countries. Tribes with no or very little contact with modern lifestyle such as uncontacted amerindians (*159*) or Hadza hunter gatherers from Tanzania (*160*) have bigger diversity of microbes thanks to a plant based diet, rich in fiber and behavioral patterns not present in industrialized countries.

The human gut microbiota evolution through life is continuous ecological succession of microbes. At the beginning pioneer microbes members of the *Proteobacteria* and *Actinobacteria* phylums colonize the intestine being nurtured by breastfeeding practices and accounting for around three-quarters of the abundances. The proportion of *Bacteroidota (Bacteroidetes)* and *Bacillota (Firmicutes)* is around 25% together at early ages, but increasing over time and decreasing again in elderly times (*161*). When solid food is introduced the community transitions from a predominantly lactic acid bacteria community to a more diverse community, where *Bacteroidota* is the phyla more represented followed by *Bacillota* (*162*). At this point the *Actinobacteria* and *Proteobacteria* phyla are in retrocess while *Bacillota* and *Bacteroidota* increase as well as other phyla (*163*). In toddlers and through aging, the *Bacillota* are increasing its presence over time. The most dominant phyla in the gut microbiome of a healthy adult is usually a *Bacillota*. Then in healthy elderly, this tendency continues being *Bacillota* the most abundant phyla with less than 25% of *Bacteroides* and other phyla. However, in super healthy elderly people that live over 100 years, centenarians, they have an increased abundance of the phylum *Actinobacteria*, summing up more than 25% of the species together with the *Bacteroides (164)*.

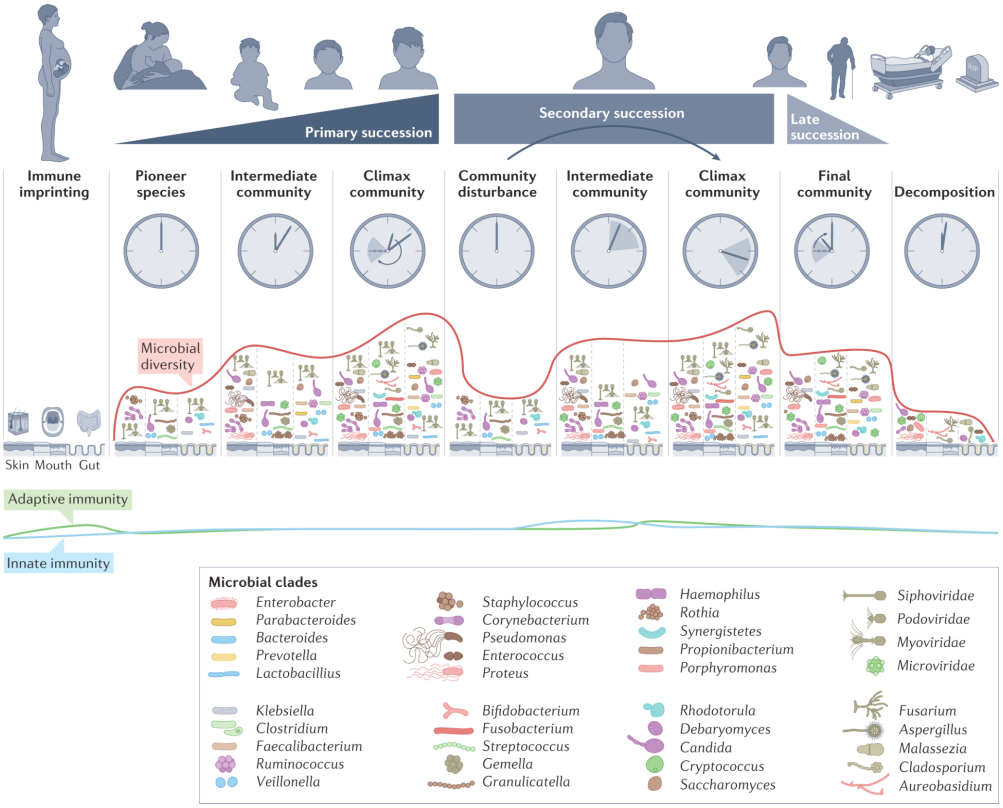


Figure 9. Ecological succession of the microbiome through life. The ecological succession in human microbiota through life. First, pioneer species colonize the baby and increase diversity in the first months of life. Then there is a continuous gain in diversity up to the climax community, that may be perturbed but is able to recover, until the late stages in life where microbial diversity decreases again. From Martino et al. (*139*).

Circadian rhythms and seasons

Daily life variations in the microbiota occur in humans and in all living organisms having an associated microbiota (*165*). Microbiota follow circadian rhythms directly or indirectly related to diet, behavior, daily hormone variations, daily changes in RNA expression or daily changes in epigenetic marks among others. The circadian rhythm clock depends on daily exposure to light and it follows daily patterns controlling all functions at the organ, tissue, cellular or molecular pathway level. The disruption of the circadian rhythm can have metabolic consequences and increase the risk of suffering metabolic and other diseases such as the metabolic syndrome (*165*), diabetes (*166*), cardiovascular diseases (*167*), mental disease (*168*) or cancer (*169*), where the microbiota might play an important role. The relationship between circadian rhythms and microbiota is bidirectional, where microbiota help to establish the circadian rhythm, and where the circadian rhythm influences the composition of the microbiota (*170*).

We have evolved for millions of years with marked seasonal variations in the availability of food and nutrients, and these variations undoubtedly shaped our symbiont microbiomes over a year, and in any animal or plant. In our westernized society, seasonal rhythms are more extensively followed by people living in small villages in the countryside of any developed country, than in big cities. But there are very few people in the world that truly follow a seasonal rhythm in terms of food and available nutrients. Hadza hunter gatherers from Tanzania are one of the few groups of people that still follow marked seasonal rhythms (Figure 10), collecting fruits and vegetables and hunting animals in a seasonal manner (*171*). Hadza live similarly as their ancestors lived thousands of years ago, having in their guts microbes that have completely disappeared in westernized societies. They had a rich plant based diet making them having an increased biodiversity and enhanced ability to digest and extract valuable nutrients from plant fibers (*160*). Uncontacted tribes from different parts of the world or societies living in less industrialized and less westernized societies, tend to live more seasonally, benefiting the microbes that evolved to seasonally changes.

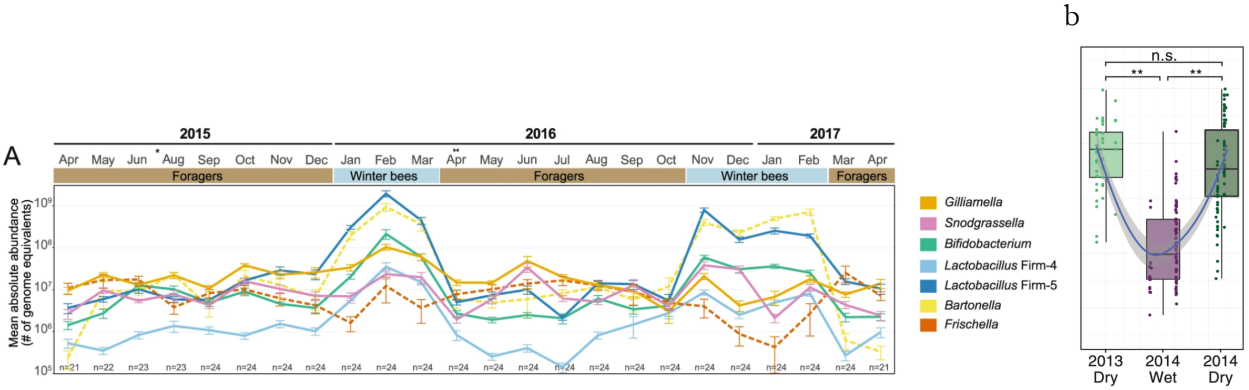


Figure 10. Seasonally microbiota composition. Seasonal gut microbiota measured in honeybees (a), extracted from Kesnerová et al (*172*). Seasonal diversity in Hadza hunter gatherers during two dry seasons and one wet season in between. From Smits et al (*171*).

Diet

Diet is one of the factors that can potentially influence the microbiota faster and in a more targeted way (*173*). People following a vegan, vegetarian or omnivore diet, have different gut microbiota compositions (*174*). Vegans and vegetarians tend to have elevated proportions of genus related to health benefits such as *Prevotella*, *Ruminococcus* or reduced proportions of *Bacteroides* (*175*), although these proportions are not always true (*176*), and depend on many factors. Omnivore diets are associated with higher proportions of *Bacteroides*, however generalizations are hard to keep valid when studying the microbiome. For example, omnivores consuming a strict Mediterranean diet, rich in fruit, legumes and vegetables, have an increase in SCFAs production, commonly associated with genera present in vegetarians and vegans (*177*). In a clinical context, knowing the enterotype (see section 3.1.4) of an unhealthy person might be a valid strategy for the stratification of this patient and personalize a dietary intervention, precisely designed to fill the empty microbial/metabolic niche to recover health. Because diet has the potential of changing gut microbiota short chain fatty acid production, emerging evidences are building the possibility that microbiota composition influences epigenetic marks in the host (*25*) (Figure 11).

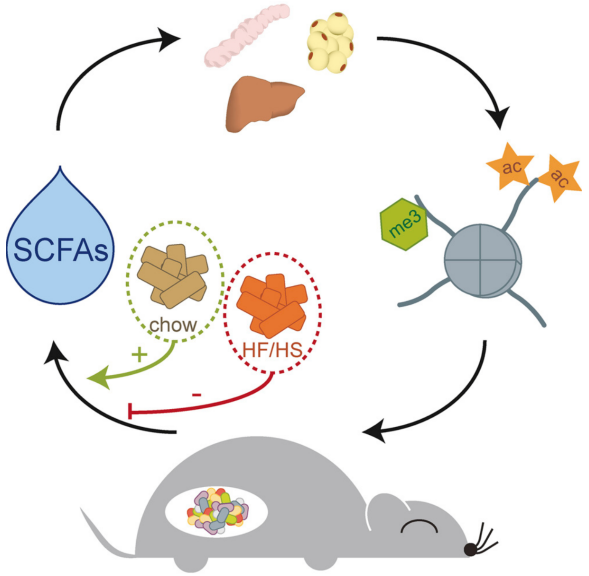


Figure 11. Microbiota produced SCFAs affect epigenetic marks. Host microbiome alters histone acetylation and methylation in multiple tissues. Western diets suppress microbiota-derived SCFAs production and chromatin effects, but when supplemented they rescue the chromatin effects. From Krautkrame et al. (*25*).

Antibiotic exposure

Antibiotic exposure is a major threat to gut microbiota diversity. In western countries, antibiotic treated meat and crops are often consumed, leading to a low-chronicle exposure to antibiotics in the gut microbiota. Western diets shift away the microbiota composition from fiber degraders to mucus consumers degrading the intestinal barrier (*178*). This chronic increase in intestinal barrier permeability is associated with a low-grade chronic inflammation (*179*). The overuse of medical antibiotics and overuse in industrial farming are the main problems in the appearance of super-resistant bacterial strains, resistant to many antibiotics at the same time and very difficult to treat during an infection

(*180*). A possible solution would be the use of bacteriophages to treat infectious diseases (*181*) in order to avoid superresistan bacterial species. The overexposure to antibiotics in the industrialized human gut microbiome is associated with a higher rate of microbial horizontal gene transfer, often promoted by consumption of processed food and exposure to low-dose antibiotics concentrations (*113*). A lower diversity often is associated with lower functionality, making antibiotic overuse a dangerous threat to human gut microbial diversity and human health in general.

Sex microbial differences

Hormones play an important role in shaping the gut microbiota composition (*182*). Different studies linked the male gut microbiota composition to a higher production of testosterone by its microbiota (*102*, *183*, *184*). In a similar way, female gut intestinal microbiota produce a higher proportion of b-glucuronidase compared to male, a necessary enzyme to conjugate estrogens in their active form (*185*). Recently, a study reported the possible retroconversion of estrogens to androgens by a bacterial species, opening the possibility to explore further the interdependencies between hormonal levels and gut microbiota composition (*186*). Sexual hormones play indeed an important role in the gut microbial composition, but there is a bidirectional chemical conversation between the gut microbiota and sexual hormones, where both partially modulate their counterpart (*187*). In relation to hormones and circadian rhythms, a recent study has shown how the murine gut microbiota is necessary for both the circadian rhythm and the maintenance of the sexual dimorphism phenotype (Figure 12)(*170*).

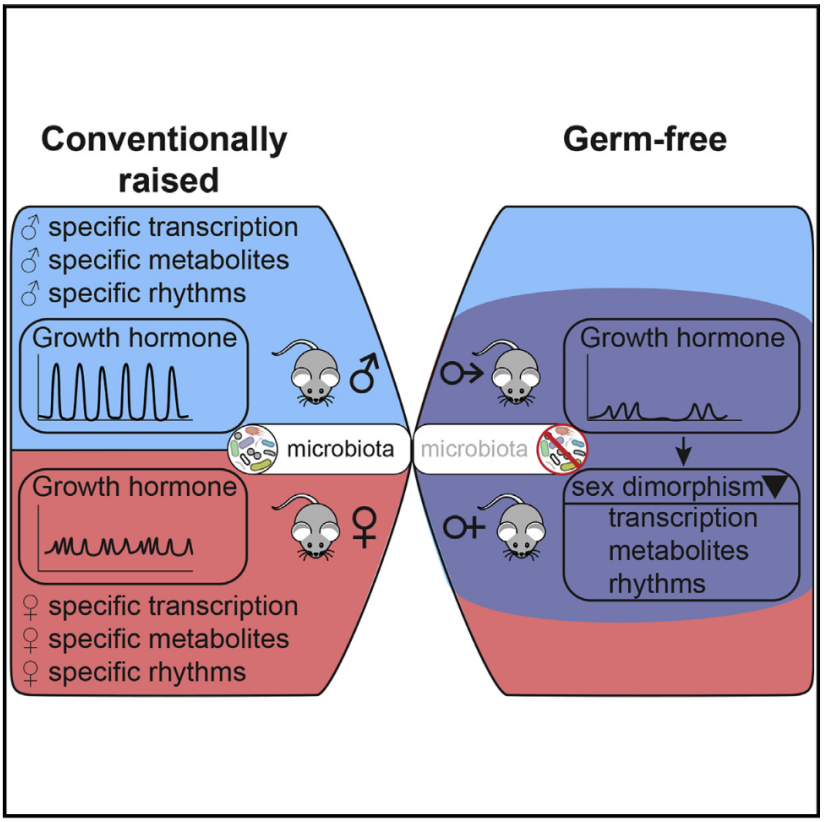


Figure 12. Microbiome, sexual dimorphism and circadian rhythms. The gut microbiota is necessary for maintaining a circadian dimorphic gene transcription profile in mice. From Weger et al (*170*).

3.2. Microbiome in human health and diseases

In the holobiont context, where we see a person and its microbiota as a unique biological entity, it is logical to observe a combinatorial effect on the incidence of diseases between the host genetics and gut microbiota composition (*188*). We know for example that aberrations in the acquisition of the initial microbiota leads to increased risks of suffering diseases. C-section newborns are at increased risk of suffering from autoimmune diseases such as asthma (*189*), obesity (*190*) or diabetes (*191*), compared to naturally delivered babies. Specific gut microbiota signatures have been found in many other diseases such as cardiovascular diseases (*192*), depression (*193*), Alzheimer (*194*), autism (*195*), colorectal cancer (*196*), breast cancer (*197*), autoimmune diseases (*198*), atherogenic diseases (*199*), also a direct relationship between microbiota and mitochondria has been observed in mitochondrial diseases (*200*, *201*). However, very few diseases have been causatively linked to the gut microbiome composition, although some specific loci of the host genome have been strongly associated with the presence of specific microbial genus (*202*). Instead, we often observe specific microbial signatures in the onset or in the disease itself, which provides space for possible dietary or microbiome modulating therapies aiming to address a particular change, missing function or overrepresentation in the microbiota of diseases suffering patients. An overview of these microbiota modulating interventions can be observed in Figure 14 at the end of this section (3.2). It has been proposed that in intense exercise athletes, a microbiota and mitochondria communication exist to support the demanding effort of this athletes (Figure 13).

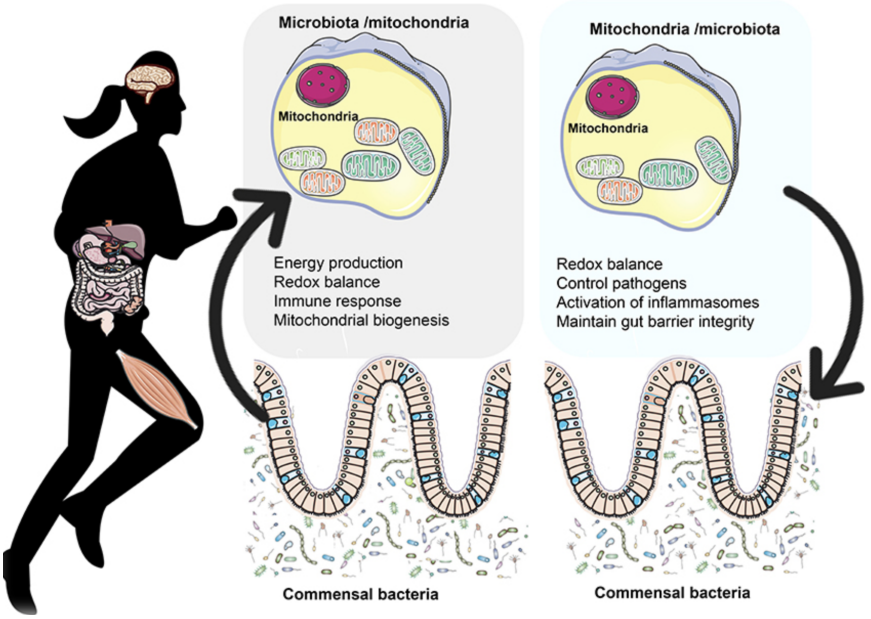


Figure 13. The mitochondria-microbiome interaction during exercise. A bidirectional crosstalk between the gut microbiota and mitochondria exists, relevant for elite athletes. Microbiota and their byproducts regulate redox balance and energy production which can affect performance, energy supplementation and motivation. From Mach et Clark (*201*).

3.2.1. **Diet interventions**

Dietary interventions are often recommended to improve gut health and health in general. A Mediterranean diet with large amounts of fruits and vegetables, and moderated ingestion of fish and meat has been proved to prevent cardiovascular risk(*203*), improve cognition(*204*) or improve DNA oxidative damage(*205*). Diet interventions can produce changes in the gut microbiota composition very rapidly and in a reproducible manner (*173*). We could see diet as the main gut microbiota modulator. We can predict the direction of the changes in the gut microbiota composition by measuring glucose response after food ingestion and having microbial 16s gut data (*206*). The consumption of higher amounts of fiber in a plant-based diet, increases beneficial lactic acid bacterial species from *Ruminococcus* and *Roseburia*, and reduces *Clostridium* and *Enterococcus* species (*175*). Then, in a context of diseases partially caused by alterations in the microbiota, we could modulate this altered microbiota in a targeted manner to recover functionality, diversity or both (*207*). Screening the microbiota first, and measuring interleukins, glucose or SCFAs levels and other biomarkers, we could then make dietary recommendations to improve health in a targeted and personalized manner.

3.2.2. **Probiotics**

Another possible treatment to address a disease's microbial signature would be the use of probiotics (life microbes) to treat a deficiency in the microbiota functionality. For example, the use of probiotics has been proved partially efficient to treat lactose intolerance (*208*), diarrhea (*209*) or in inflammatory bowel diseases (*210*). For example, *Bifidobacterium lactis* has been used to specifically inhibit the toxic effect of wheat gliadin (*211*). But, although some progress has been made in the use of probiotics to treat different diseases, the challenge remains the engraftment of the probiotic in the intestinal tract, this process depending on the resident microbiota, a part from the probiotic itself (*212*). To achieve a beneficial long-term symbiosis between the host and the probiotic strain, the engraftment must occur to avoid continuous ingestion of the probiotic. For that, using an ecological framework is necessary to interpret, design and predict the impact of a given probiotic strain or bacterial consortia (*213*).

3.2.3. **Prebiotics**

Probiotics supply life microbes to increase the diversity of the gut microbiota, but another strategy to achieve similar results is the use of prebiotics. Prebiotics are any kind of polysaccharide that the host cannot digest by himself and it is a useful source of energy for the resident microbiota. Microbiota have many functions encoded in their genetic material that allows the digestion of many polysaccharides that pass through the host intestine without being digested by host enzymes. These polysaccharides or fibers can be used to modulate the intestinal microbiota and enhance the metabolic functions that the resident microbiota can perform, usually benefiting the host (*214*). Prebiotics hold a great potential to treat intestinal disorders linked to many diseases, however a limited number of trials and studies have been conducted to assess the feasibility of using microbiota modulating prebiotics to treat specific functional gut disorders (*215*, *216*). A combination of pro- and prebiotics might give an advantage to successfully modulate the microbiota in the wanted direction.

3.2.4. **Postbiotics and symbiotics**

Postbiotics are any non-living microbe (i.e. spores) or some of their parts that can be beneficial for the host. Salminen et al. defined the term as follows: "a preparation of inanimate microorganisms and/or their components that confers a health benefit on the host" (*217*).
Symbiotics are life microorganisms and their by-products that are selectively used by other host microorganisms that can be beneficial for the host. One can say that symbiotics are the combination of probiotics and prebiotics together. Swanson et al defined the term as follows: "a mixture comprising live microorganisms and substrate(s) selectively utilized by host microorganisms that confers a health benefit on the host" (*218*).
Using symbiotics and postbiotics in basic research, translational experiments or clinical trials, can help researchers to modulate the gut microbiota, prevent or treat specific diseases. Postbiotics have been demonstrated to have a beneficial effect on the host, comparable to probiotic supplementation (*219*). Synbiotics are more often used because of the increased potential benefit of supplementing with a pre- and probiotic at the same time. Supplementing a single specific bacterial strain and a single prebiotic molecule, in a research lab setting with mice, have demonstrated beneficial and microbiota modulating effects to the treated mice (*220*).

3.2.5. **Fecal microbiota transplantation (FMT)**

Another, more drastic microbiota modulation strategy is the transfer of the microbiota itself, called a fecal microbiota transplantation or FMT. FMT consists in replacing the resident microbiota with a new (healthier) microbiota using a colonic enema or through oral capsules. Patients suffering from recurrent *Clostridium difficile* infection (RCDI) recover and prevent RCDI after the first fecal microbiota transplantation in 96.2 % of the times (*221*). This figure augmented the odds of using FMTs for treating other diseases such as obesity (*222*), depression (*223*), Chron's disease (*224*), autism (*225*) or melanoma (*226*) among other disease potentially treated by FMTs.

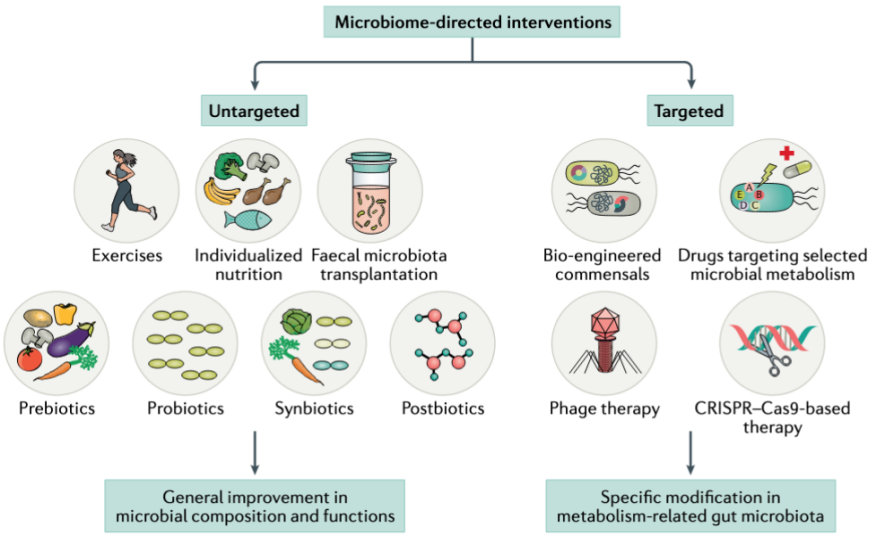


Figure 14. Microbiota-modulating interventions. Targeted and untargeted microbiome-directed interventions in humans with disrupted gut microbiomes. From Fan and Pedersen (*227*)

3.3. Microbiota-host axes

Lately, many microbiome-host axes have been proposed to highlight often a tissue or organ-microbiota specific relationship. Some of the most studied and well characterized axes are the microbiota gut-brain axis, gut-liver axis, gut-lung axis, or microbial gut-metabolism axis(*103*, *228–230*).

3.3.1. Microbiota gut-brain axis

The microbiota gut-brain axis is one of the hot-topics in microbiome research due to the possibility that gut microbes can affect the host behavior (*103*). A large portion of the nervous system is present in the gastrointestinal tract, and in close contact with microbes in the intestines, which produce molecules functioning as neurotransmitters for the host (*31*), hormones (*231*), and substrates for brain metabolism (*232*) (Figure 15).

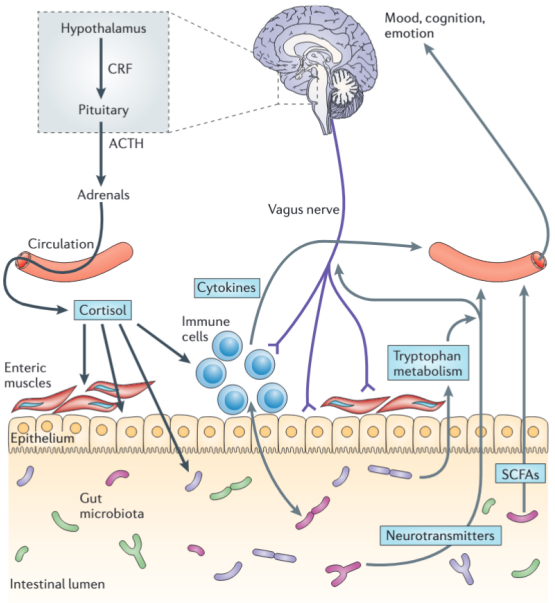


Figure 15. Metabolic pathways involved in gut-brain communication. Microbiota can modulate the gut-brain axis by synthesizing, consuming or transforming dietary products, by affecting the endocrine, immune or neural metabolism. From Cryan and Dinan (*35*).

3.3.2. Microbiota gut-liver axis

The gut-liver axis is also of special interest due to the metabolic function of both the liver and the gut microbiota (*228*) Understanding how the gut microbiota communicates to the host liver, and vice versa, has shown bile acid metabolism as a key bidirectional metabolism communication (*233*). Newly discovered secondary bile acids are conjugated by gut microbes, enriched in disease states and associated with microbiome dysbiosis (*34*).

3.3.3. **Microbiota gut-immune axis**

Together with the enteric nervous system, the immune system is of special interest because of the potential role of the microbiota modulating nervous system immune cells (*234*, *235*). It has been proposed that an occidental life-style and diet with antibiotics overuse has selected a less resilient and less diverse microbiota needed for a balanced immune response (*229*), which may explain the rise of autoimmune and inflammatory diseases in these westernized countries.

3.3.4. **Microbiota gut-metabolism axis**

Gut microbiomes have the potential of producing thousands of chemicals, with more chemical diversity encoded in their genomes than the one encoded in the host genome (*230*). This fact puts the microbiome in a position of importance in the metabolic homeostasis of the host. Many diseases, or almost all of them, have a metabolic component that can be affected/controlled by gut microbiota metabolism. Inflammatory bowel diseases (IBD), cancer, diabetes, obesity, cardiovascular dyslipidemia or metabolic endotoxemia have important metabolic alterations that are linked to the gut microbiome metabolism (*236*).

4. Epigenetics

Conrad Hal Waddington coined the term "epigenetics" in the early 1940's, presenting the concept in the first ever epigenetic publication "*Canalization of development and the inheritance of acquired characters*" where Waddington proposed a mechanism for explaining inheritance of what we know today as epigenetic mechanisms (*237*). Since then, many technological advances have happened and many biological and epigenetic processes have been uncovered, but Waddington's message is still valid today. Epigenetics focus on the study of inherited phenotypic changes that do not involve alterations in the DNA sequence. The prefix "epi-" in Greek means "over", "outside of" or "around", meaning that epigenetic events are features "on top of" or "around" the classical genetic view of inheritance. These features that change the phenotype without altering the DNA sequence are reversible and influenced by environmental stimuli. Therefore, susceptible to be modified through a drug, diet or microbiome intervention.
Improvement of sequencing technologies and molecular tools in the past twenty years have moved forward the epigenetic field, from barely being able to sequence the whole genome in 2001(*238*) to being able to perform a whole genome bisulfite sequencing (WGBS) experiment to measure the degree of methylation in a DNA sample (*239*), or to capturing the long range 3D interactions between different regions of the genome with the Hi-C method (*240*), or even to edit the epigenome using CRISPR gene editing technology (*241*). New epigenetic knowledge and molecular techniques are created every year to characterise and define different epigenetic aspects and advance the field towards a mechanistic understanding to treat and prevent diseases.

There are at least five levels of epigenetic regulation events involving different mechanisms and molecules that can modify and control gene expression in a cell. The first, and most studied epigenetic feature is DNA methylation, which implies the addition of a methyl group in a cytosine of a CpG dinucleotide (*239*). Another well studied epigenetic features are histone tail modifications, which consist in the addition of chemical modifications such as methylation or acetylation (among many others) to the histone tails in the core histones of a nucleosome (*242*, *243*) that are capable of regulating gene expression. RNA modifications are another level of epigenetic features controlling protein translation with functional repercussions discovered recently (*83*). Non coding RNA regulation of gene expression is a complex epigenetic feature that can involve different types of non-coding RNAs such as miRNAs, snRNAs, piRNAs or lncRNAs (*242*). An epigenetic gene regulation can be due to the 3D structure of the genomic material, where two sequences can be megabases away from each other in the linear sequence, but very close in a 3D space. These epigenetics events preserve memory of past states and signals, and activate future behavior in the absence of both the initial signal and alterations in the DNA sequence (*244*). There is a complex crosstalk and different interactions between these five epigenetic features at many levels and are mainly not created, deposited or recognized in isolation, they form part of a complex and interrelated collection of cell features (DNA and RNA marks, chromatin conformations, histone modifications, metabolism…) that regulate cell fate.
In this section (4) I review most common epigenetic strategies used by cells and organisms to regulate gene expression and transcription. From covalent modifications in the DNA, RNA and histones tails, to regulation of gene transcription by non-coding RNA and genome 3D architecture.

4.1. DNA modifications

DNA is composed of four main bases cytosine (C), guanine (G), thymine (T) and adenine (A), where each of them can be theoretically chemically modified (*245*). However, 5-methylcytosine (5-mC) and 5-hydroxymethylcytosine (5-hmC) are the most stable and better characterized modifications in the DNA molecule (Figure 16). The discovery of the bisulfite conversion (*246*, *247*) and the invention of the methyl specific PCR, (*248*) put the bases for WGBS, and made possible the development of other applications such as the oxidative bisulfite sequencing to map at the same time 5-mC and 5-hmC (*249*). We can distinguish in the DNA molecule different parts or locations that play different roles in the regulation of genes. Promoters and their methylation levels are known to be responsible for the regulation of gene expression (*250*), enhancers are other distant regulatory regions susceptible to regulating gene expression dependent on methylation levels (*251*). But the levels of methylation in introns, exons or untranslated regions can also influence gene regulation (*252*).

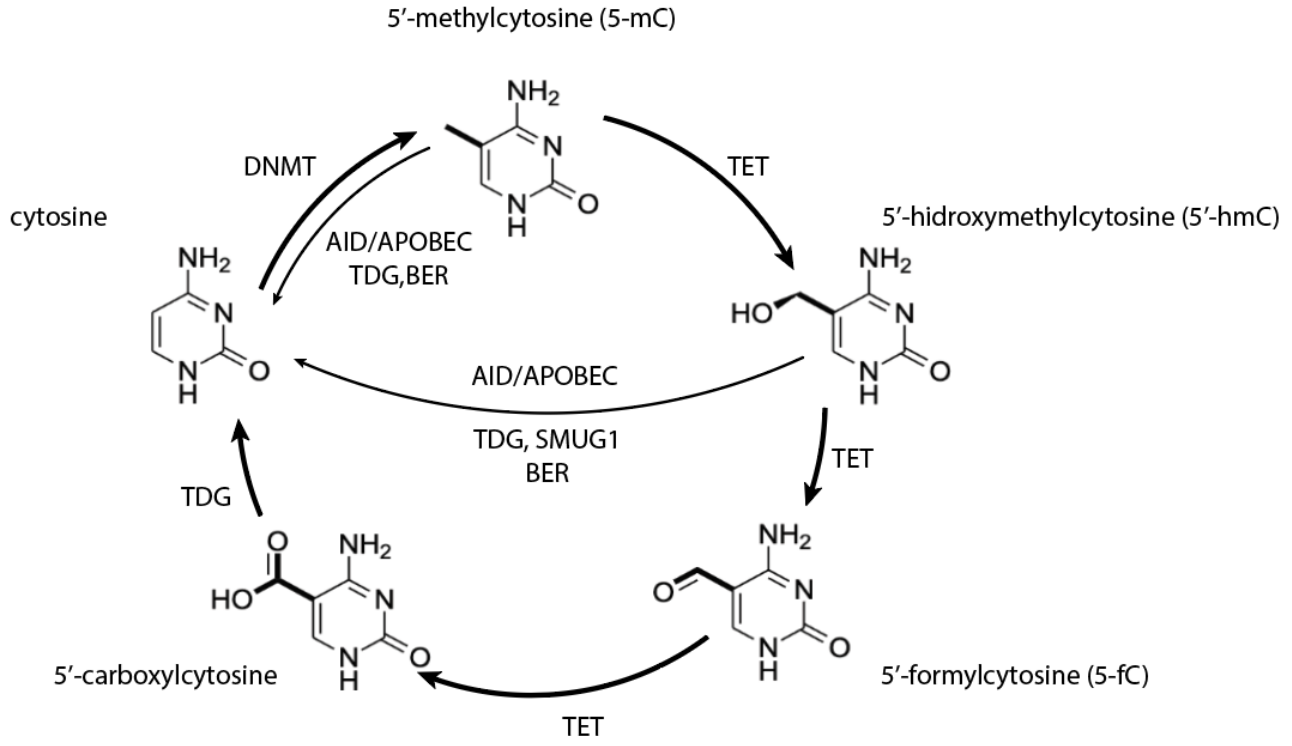


Figure 16. Cytosine and modified cytosine pathway. DNA base cytosine can be enzymatically modified and converted to 5'-methylcytosine (5-mC), 5'-hydroxymethylcytosine (5-hmC), 5'-formylcytosine (5-fC) or 5'-carboxylcytosine (5-caC) by different enzymes shown in this figure. These modifications are believed to exert different epigenetic effects on the DNA transcription machinery. TET: ten eleven translocation enzymes; DNMT: DNA methyltransferase; TDG: , thymine DNA glycosylase; BER: base excision repair complex; SMUG1: strand-selective monofunctional uracil-DNA glycosylase 1; AID: activation-induced cytidine deaminase; APOBEC: apolipoprotein B mRNA editing enzyme, catalytic polypeptide. Adapted from Branco et al. (*253*).

4.1.1. **5' methylcytosine (5-mC)**

DNA methylation is the most studied epigenetic feature in biomedical research. DNA methyltransferases (DNMT1, DNMT3a and DNMT3b) are responsible to transfer a methyl group from SAM to the fifth carbon of the cytosine in the DNA molecule (*254*). Since its discovery in bacteria in 1925 (*255*), DNA methylation has been investigated in many biological contexts and experiments (*256*). The main functions of methylated DNA are to repress gene expression, control embryonic development, controlling enhancers functions, transcription elongation or alternative splicing among others (*257*). To study and map 5-mC in the genome, researchers often use bisulfite sequencing (BS-seq = WGBS) (*258*) or reduced representation bisulfite sequencing (RRBS) (*259*).

DNA methylation is a dynamic rather than a fixed epigenetic mark, and changes in the methylation status of specific cytosines happen over a lifetime of a cell in response to environmental stimuli. Demethylation happens when the methyl group of a cytosine is removed by oxidation and deamination processes regulated by different enzymes and triggered by changes in environmental stimuli (*260*). The above cytosine modifications represent the different states of the demethylation process in a cytosine of the DNA molecule (Figure 16).

4.1.2. **5'-hydroxymethylcytosine (5-hmC)**

DNA hydroxymethylation is an intermediate state of the removal of the methylation group from the carbone 5 in a cytosine. This process is mediated by ten-eleven translocation enzymes or TETs (TET1, TET2 and TET3) which oxidize the methyl group with the addition of a hydroxyl group (*261*). The function of 5-hmC is not clear and might be only an intermediate between methylated and unmethylated cytosine, being this its main cellular role (*253*).

4.1.3. **5'-formylcytosine (5-fC)**

Another step forward in the DNA demethylation process is the conversion from 5-hmC to 5-formylcytosine (5-fC). This process is also mediated by TET enzymes and with the help of mammalian DNA glycosylase TDG, mediates the demethylation process. Mapping of 5-fC in the genome has revealed that 5-fC preferentially occurs at poised enhancer, a distinct set of distal regulatory elements, and together with p300, they are remodeling epigenetic states of enhancers (*262*).

4.1.4. **5'-carboxylcytosine (5-caC)**

After the oxidation to 5-fC, TETs can further oxidize this cytosine species to 5-carboxycytosine (5-caC) which then can be further demethylated to cytosine by the TDG DNA glycosylase. It has been demonstrated that the main function 5-caC together with 5-fC is to slow down and retard RNA polymerase II (Pol II) elongation on gene bodies (*263*). To map the 5-caC in the genome Lu et al. have developed the chemical modification-assisted bisulfite sequencing method (CAB-seq) (*264*).

4.1.5. **Other DNA modifications**

Adenine modifications, such as N6-methyl-2'-deoxyadenosine have been observed in the DNA of prokaryotes and eukaryotes. However, the function of this modified DNA base is still debated, it has been suggested a role in modulating DNA structure and transcription (*265*).
A thymine modification has also been observed and can be detected taking advantage of the oxidation of thymine to 5-hydroxymethyluracil (*266*). Guanine oxidation to 8-oxo-7,8-hydroxyguanine might be another modification capable of regulating gene expression (*267*). But these marks in other bases different than in C, are much less studied than the canonical 5-methylcytosine in the DNA molecule.

4.2. Histone PTMs

The nucleosome is the basic unit of the chromatin for packaging the DNA in the nucleus of eukaryotic cells (*268*). Nucleosomes are linked between them by a linker DNA, and the histone H1 organizes nucleosomes into the chromatin superstructure. Each nucleosome contains 146 bp of DNA wrapping eight core histone proteins, two copies of each of the core histones H2A, H2B, H3 and H4. Nucleosomes and chromatin are responsible for allowing the DNA to be highly compacted, less accessible, or more relaxed and easily regulated inside the nucleus. Each of the core histones in the nucleosome has a tail protruding from the nucleosome core. These tails serve to recruit transcription factors and regulate 3D conformation of the chromatin, allowing or preventing access to the DNA by the transcription machinery (*269*). Histone tails can be highly modified by enzymes, and depending on the type of modification, chromatin will be more or less accessible for transcription factors and polymerases, changing its 3D conformation. To date, many modifications have been described from what is called the "histone code" (*270–272*). Most studied modifications are acetylation and methylation. In general, acetylation of the histone tails is associated with open chromatin and active transcription, while methylation of histone tails is associated with more closed chromatin and transcription repression. The two principal techniques to quantify histone modifications are mass spectrometry (*273*, *274*) and chromatin immunoprecipitation (ChIP) and ChIP-seq (*275*, *276*). Regulation of chromatin accessibility to transcription factors, polymerases and other assistant proteins, depends on the residue, position and type of modification in the histone tail. And often will determine its activation or repression function (*277*). It is suggested that an extensive crosstalk exists between acetylation and other histone PTMs, to coordinate and modulate chromatin functions (*278*). These different histone marks (Figure 17) are not exerting their function separately, they act synergically or antagonistically, and form part of a complex and interrelated histone language that regulates gene expression (*279*).

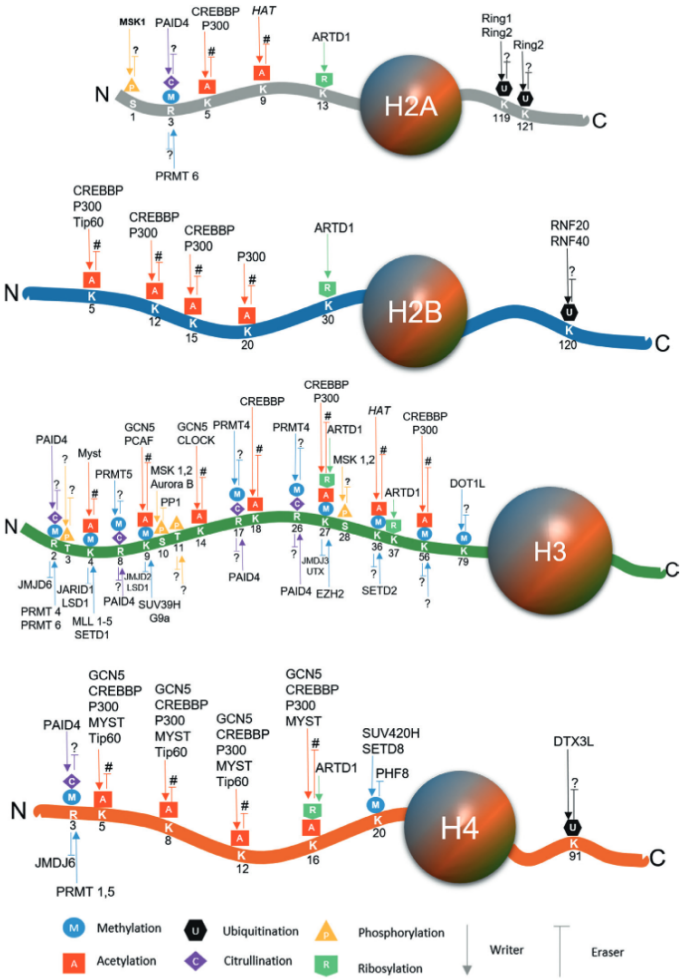


Figure 17. Histone PTMs altered in cancer. Individual core histones, H2A, H2B, H3, and H4 with their N-and C-terminal tail, different histone PTMs, and their respective modifiers. From Sharda et al. (*280*)

4.2.1. **Acetylation in histones**

Histone tail acetylation is associated with open chromatin and active gene transcription (*281*), while the lack of acetylation tends to correlate with less gene transcription (*282*). Histone acetylation has been detected on lysine and arginine residues of H3 and H4 histone tails (*283*). The acetyl group is deposed to lysine residues by histone acetyltransferases (HAT) that catalyze the group transfer from the donor acetyl coenzyme A (ac-CoA) to the receptor lysine residues in the histone tails (*284*, *285*). Histone deacetylases (HDAC) do the opposite reaction, removing the acetyl group from lysine residues and transferring it to ADP-ribose (*283*). Sirtuins are NAD-dependent lysine deacetylases that remove the acetyl group to regulate gene transcription (*286*). For example, SIRT1 controls the cellular response to stress by regulating FOXO forkhead transcription factors, genes believed to be responsible for organism longevity (*287*).

4.2.2. **Methylation in histones**

Histone tail methylation occurs on lysine and arginine residues and is a dynamic or stable molecular process in different stages of a cell life (*288*, *289*). Lysine methylation and lysine demethylation have been both associated with gene-activation and gene-repression in bivalent marks, and ensuring appropriate gene expression in monovalent marks in cell differentiation (*290*). It has been associated with activation and repression of transcription depending on the position and type of arginine in the histone tail (*277*, *291*). There are many methyltransferases that are responsible in depositing a methyl group in specific residues of histone tails, for example there are four enzymes that generate H3K36me2 (NSD1, NSD2, NSD3 and ASH1L), two for H3K9me1 (PRDM3 and PRDM16), one for H4K12 (KMT9), among many other lysine modificators (*292–294*). H3K4 demethylation is mediated by KDM5, a Y-linked gene that encodes an histone demethylase in a sex-specific manner (*295*). Different KDMs play major roles at specific stages in the cell cycle, regulating processes such as the formation of replication origins, repression of cell cycle regulator p21 or regulating chromosome segregation (*296*).

4.2.3. **Phosphorylation in histones**

Histone tail phosphorylation has been observed in the context of DNA damage, when phosphorylated histones delimit large chromatin regions around the DNA break (*297*). Other functions of histone phosphorylation are for example the phosphorylation in tyrosine 41 and threonine 45 residues, implicated in transcription and replication, respectively (*298*). These modifications are located near where the DNA enters-exit the nucleosome, significantly increasing the unwrapping and accessibility of transcription regulatory complexes when phosphorylated (*299*). Phosphotransferases, also called kinases, are responsible to transfer the phosphorus group from ATP to serine, tyrosine or threonine(*300*), and are removed by phosphatases (*301*). The addition or removal of a phosphorus group in histones is associated to activation and repression of gene expression, respectively, and present in many developmental, cell cycle, tumor formation or response to environmental stresses processes (*297*).

4.2.4. **Propionylation, butyrylation and crotonylation in histones**

Histone tails can be propionylated, butyrylated and crotonylated. Lysine H3K9 butyrylation is negatively regulated by high fat diet and stress, reducing H3K9 butyrylation levels in promoters under oxidative stress and diminishing gene expression (*302*). Propionylation of lysines in histone tails is mediated by p300 acyltransferase and requires propionyl-CoA to transfer the propionyl group to the histone tails (*303*). Sirtuins catalyze the depropionylation of histones in a NAD+ dependent manner and the propionyl group is transferred to ADP-ribose (*304*). The size or length of the acyl chain added to the lysine residues seems to be inversely correlated with the activity of HAT, the smaller the chain the more HATs activity (*303*). Debutyrylation and depropionylation are mediated by class III HDAC such as SIRT2 deacetylase (*283*). Histone propionylation, similarly to other acyl modifications, modulates gene transcription depending on position and residue modification (*305*). Histone crotonylation has also been observed in histone tails and positively correlated to gene expression (*306*). Again, p300 acyltransferase is one of the responsible enzymes to deposit crotonyl groups on histone

lysines depending on the abundance of crotonyl-CoA (*307*). SIRT1, SIRT2 and SIRT3 are able to catalyze the hydrolysis of lysine crotonylation on histones and other proteins (*308*).

4.2.5. **Other histone modifications in histones**

Many modifications have been described in different residues of histone tails (Figure 17) (*272*). Glycosylation on histone tails by O-GlcNAcylation affects chromatin remodeling and gene expression. Depending on the position, O-GlcNAcylation can activate or repress gene expression by modulating the activity of methyltransferases and TET enzymes, promoting histone methylation or demethylation (*309*). Macrophage histones are ADP-ribosylated at transcriptionally active regions, increasing the accessibility to promoters and increasing gene transcription (*310*). Histone lysine lactylation has recently been described as a novel histone mark, capable of activating gene expression in bacterially challenged macrophages, showing how metabolism can directly regulate gene expression (*80*). Histone lysine succinylation and manolylation have been proven to modify gene expression and link metabolism to chromatin structure and accessibility to nucleosome DNA (*311*, *312*). Ubiquitination of histones has many functions, from signaling where to repair DNA breaks (*313*, *314*), to regulating DNA methylation of specific genes (*315*), or being associated with highly expressed genes (*316*).
In summary, cells use histone modifications to fine tune the recruitment of transcription and replication factors, helping the cell to control gene expression with a diverse set of molecular mechanisms.

4.3. RNA modifications and transcription regulation

RNA modifications and transcription regulation events such as alternative splicing of a gene or the regulation of transcription by transcription factors, are also considered epigenetic events. RNA modifications have been long described and recent advances in sequencing technologies allowed a better understanding of their role in cell fate (*317*). Alternative splicing is a major source of protein diversity regulated by histone PTMs and chromatin structure, those determining which parts of the genome are expressed and how these parts are spliced (*318*). Transcription factor (TF) assists other transcriptional machinery to regulate gene expression by binding to the DNA molecule and helping other factors with the transcription (*319*).

4.3.1. **RNA modifications**

N^6-methyladenosine (m6A) has recently been the focus of many research projects for its widespread regulatory mechanism controlling gene expression in many biological processes. It has been connected to cell differentiation, cancer progression and regulation of gene translation (*317*). Mapping the m6A into the transcriptome has revealed a transcriptome-wide enrichment of this mark on untranslated regions (UTRs) and near stopping codons (*320*, *321*). The methodology to map the positioning of m6A into the transcriptome is called m6A-seq and has permitted scientists to accurately map this RNA modification (*321*). RNA cytosine methylation on transfer-RNA (tRNA), promotes protein synthesis during development (*322*), and has been linked to leukemia development (*323*).

Pseudouridine is the most abundant non-coding RNA post-transcriptional modification, stabilizes the RNA molecule and for example, the messenger RNA (mRNA) has hundreds of pseudouridine sites. The majority of pseudouridine sites in mRNA appear in response to nutrient starvation, connecting the metabolism and nutrient availability to an epigenetic mechanism controlling gene expression (*324*). Pseudouridine synthases have been shown to regulate alternative splicing of pre-mRNA controlling gene expression in a tissue-specific manner (*325*). The conversion of adenosine to inosine is another post-transcriptional modification that occurs in mRNA and non-coding RNA. Inosine properties are similar to guanine and therefore, this post-transcriptional modification can alter codons, remove or introduce splice sites and affect the base pairing of the RNA (*326*). RNA post-transcriptional modifications are another level of epigenetic regulation, controlling the protein production and RNA interaction with other RNAs and DNA.

4.3.2. **Alternative splicing**

Splicing and alternative splicing can be considered another epigenetic mechanism of controlling protein translation from RNA, those affecting the phenotype. Splicing is the process by which the introns of a gene are excluded from the messenger RNA molecule, and exons joined together to form a mature mRNA. Alternative splicing is the process by which the exons are joined in different combinations, translated into different but related protein isoforms with different functionalities, being a major source of protein diversity. Histone modifications and chromatin structure have been pointed to have a regulatory function in alternative splicing, contributing to RNA processing (*318*). Alternative splicing is responsible for creating the diverse and complex proteome of multicellular organisms by generating different protein isoforms from a relatively small set of genes (*327*). There are five major alternative splicing events occurring in eukaryotic cells that control the protein isoform and phenotype in the cell: mutually exclusive exons, cassette alternative exons, alternative 3' sites, alternative 5' sites and intron retention events (*328*, *329*). Some of these mechanisms are perturbed in cancer and constitute a hallmark of cancer (*330*).

4.3.3. **Transcription factors**

Transcription factors are proteins that assist other proteins in the transcription process by either binding directly to the DNA molecule or interacting with other proteins of the transcription machinery. For example, some transcription factors have the dual ability of binding specifically to a DNA sequence and interact with enzymes, such as DNA methyltransferases (DNMT), that can promote specific DNA methylation patterns at promoter regions (*331*). Conservation between the RNA Polymerase I, II, and III transcription initiation machineries has shown to direct the initiation of RNA synthesis (*332*).

4.4. Non-coding RNA

Non-coding RNAs is a diverse molecular field in terms of functionality and structure, and is considered another layer of epigenetic regulation. Some of the molecular functions driving health or diseases involve the non-coding RNA world. Endogenous small-interference RNA (siRNA), micro RNA (miRNA), piwi-interacting RNA (piRNA) or long non coding RNAs (lncRNA) are some of the

non-coding RNA that are transcribed, not translated but have important regulatory functions in gene expression and cell homeostasis (*333*).

4.4.1. **lncRNA**

Long non-coding RNAs are molecules between 1k and 10k residues in length that are usually transcribed by POL II and are polyadenylated (*333*). lncRNAs can be classified in five major categories: sense, antisense, bidirectional, intronic or intergenic, and they have different functions and evolutionary origins (*334*). Their functions goes from inactivating the X chromosome in females by the Xist lncRNA gene (*335*), to changing the cellular metabolism to promote viral replication by the lncRNA-ACOD1 gene (*336*), or controlling the recruitment of polycomb repressive complex II (PRC2), a major epigenetic modulator, by the intronic lncRNA ANRASSF1 gene (*337*). The lncRNA can act in *cis* or *trans* positioning, by acting on the same strand (*cis*) or on the opposite strand (*trans*) of the targeted loci. And lncRNA are suggested to offer, sometimes, an advantage over protein epigenetic regulation (*338*).

4.4.2. **miRNA**

Micro RNAs (miRNA) are also considered another mechanism of epigenetic regulation. miRNAs are single stranded small RNA molecules, 18-25 nucleotides in length and have epigenetic regulatory function in gene expression (*339*). miRNAs exert their function by specifically binding to its complementary region in the pre-mRNA molecule, and it can induce transcription when interacting with the promoter region (*340*), or repress transcription when binding to the 5'UTR or coding regions (*341*, *342*). They activate or inactivate translation into proteins of the targeted RNA, depending on the cellular context and where they bind (*343*, *344*). Or they alter chromatin states by triggering an enhancer-mediated mechanism (*345*). This class of non-coding RNAs have a diverse mode of action, from regulating gene expression to, and in a completely different context, modulate host-microbe interactions by modulating microbiota composition when excreted into the intestinal lumen (Figure 18) (*346*).

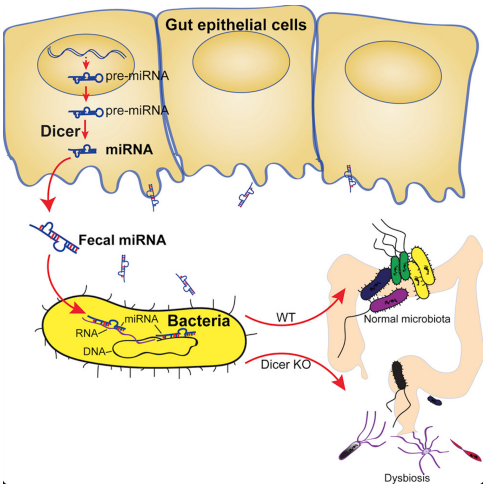


Figure 18. Host miRNA control of microbiota. Enteral epithelial cells release miRNA to control the abundance and composition of bacterial species in the gut lumen. From Liu et al (*346*).

Other small non-coding RNAs: piRNA, snoRNA, siRNA

Piwi-interacting RNAs (piRNAs) are another type of small non-coding RNAs characterized by having a length of 21-35 nucleotides in length, silencing transposable elements and controlling gene expression and fighting viral infections (*347*). For example, piRNA silences transposons by inducing methylation on the transposable elements (*348*). Small nucleolar RNAs (snoRNA), process ribosomal RNA (rRNA) but they are also know to function as guide RNAs in the post-transcriptionally modification to pseudouridine, 2'-O-methylation of nucleotides in a variety of cellular RNAs (*349*). There are two major classes of snoRNAs depending on the conserved sequence, and exert a variety of functions, from regulating alternative splicing and posttranscriptional RNA modifications, to being the source of other short regulatory RNAs (*350*). Small interference RNAs (siRNA), are a category of double strand small RNAs (20-24 bp) that interact with targeted regions and usually silence their expression by degrading mRNA after transcription. The formation of the RNA-induced silencing complex (RISC) uses siRNA and other small RNAs together with Argonaute proteins to form the silencing complex (*351*).

4.5. Genome architecture

How the genomic sequences are compacted, folded, distributed and dynamically interacting between them in the three-dimensional space inside the nucleus, constitute another degree of epigenetic control of gene expression. Chromosome interaction maps show that inactive domains are condensed in space, whether active domains can form intra- and extra-chromosomal interactions, reaching out of the chromosome domain and forming functional domains (*352*). We can classify the chromatin interactions in four different levels: nucleosome-nucleosome, chromatin loops, topologically associated domains (TADs) and compartmentalization of megabase-scale chromatin (Figure19) (*353*). Where the hierarchical clustering of chromatin elements determine the accessibility to DNA by transcription factors and transcription complexes, those controlling gene expression.

Cis-regulatory elements such as promoters, enhancers and super-enhancers are distant sequences that regulate a targeted gene sequence thanks to the linear proximity in the case of promoters, or 3D folding structure of the chromatin in the case of enhancers and super-enhancers, bringing the regulatory region and the targeted gene closer to each other in the 3D space. Cis-regulatory elements have their targeted genes in the same strand of the DNA and can regulate more than one gene. Experimental confirmation has demonstrated physical contact between the regulatory elements and gene promoters in the activation of the targeted genes by chromatin looping (*354*). For example, in type 2 diabetes (T2D), pancreatic islets genome architecture and regulatory variants impact insulin secretion in T2D patients (*355*), or hypomethylation of specific enhancers in the DSCAML1 gene is associated with upregulation of the BACE1 and increase of amyloid plaques and cognitive decline in Alzheimer diseases (*356*). Cis-regulatory elements are another gene expression regulatory layer that depends on the 3D structure of the DNA and chromatin. While cis-regulatory elements are made by DNA elements that encode transcription factor (TF) recognition sequences, the trans-regulatory elements are those TF that recognise DNA sequences and regulate transcription. Mammalian evolutionary selection on gene regulation elements has favored the conservation of trans-regulatory elements allowing a major plasticity of cis-regulatory elements (*357*).

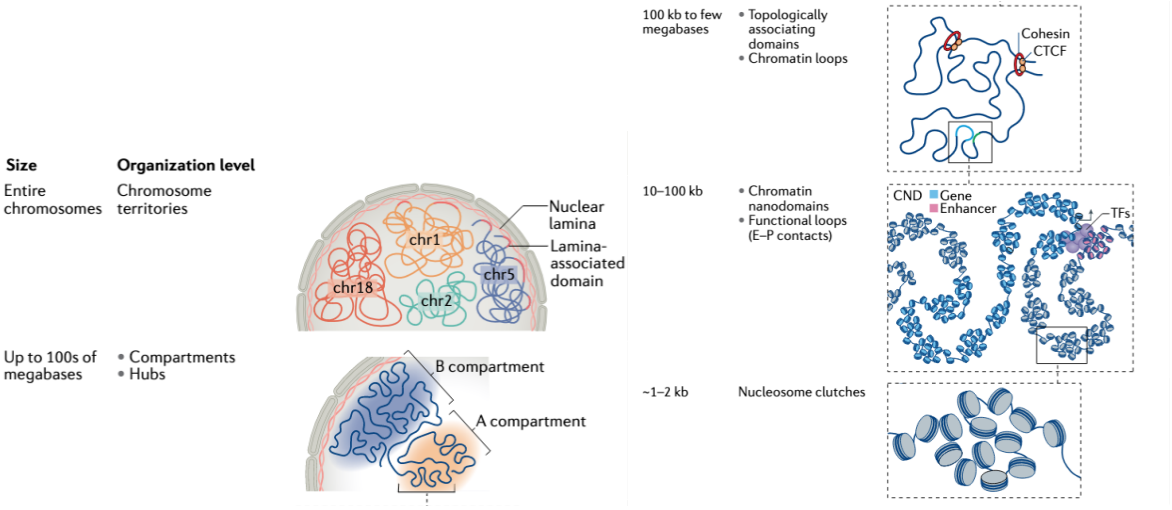


Figure 19. 3D genome architecture. Different levels of organization exist within the genome. The biggest level of organization is chromosome, then chromosome territories, compartments and hubs. Then topologically associated domains and chromatin loops. Then functional loops and nanodomains and finally nucleosome clusters. From Jerkovic and Cavalli (*358*).

5. Multi-omics sciences

The technological advances in mass spectrometry (MS) and next generation sequencing (NGS) of the past 20 years has led us today to have a powerful set of technologies capable of measuring different molecular layers at unprecedented resolution to understand the biological bases of life in a cell, tissue or holobiont. We can for example measure the small molecules forming the metabolome of a tissue, or measure the protein levels in our model system by MS (*359*, *360*); or measure the transcription levels in a tissue (*361*), or measure if a specific transcription factor is binding to a specific sequence of the DNA molecule in the nucleus (*362*), or measuring the chromatin accessibility, the methylation levels in the DNA and the transcriptome in a single cell at the same time by NGS techniques (*363*). We can study our samples' biological properties using MS and NGS technologies to generate the data, bioinformatic tools to process and analyze the data and extract meaningful biological information. We refer to omics sciences at any MS and NGS experiment that can be conducted at high throughput and generate high dimensional datasets.

To characterize the biological molecules present in the mouse holobiont (host + microbiota), we have used 4 major omics in this PhD research: metabolomics, proteomics, genomics (including transcriptomics and epigenomics) and metagenomics. Although there exist other omics fields such as lipidomics, epitranscriptomics or metaproteomics, which measure lipids, RNA epigenetic marks or proteins in bacterial communities respectively. In this PhD work, I have studied the mouse holobiont measuring metabolites, metals and histone PTMs using MS techniques, and the microbiota composition, liver transcriptional profile and liver DNA methylation levels using NGS techniques.
In this section (5) I introduce the 2 main technologies used for the 4 major omic sciences I used in this thesis, I define and summarize key concepts and some technical aspects of MS and NGS omics, and briefly describe the statistics used to integrate the different omics to study the mouse holobiont.

5.1. Mass spectrometry in omics science

Mass spectrometry (MS) is one of the technologies that has improved dramatically in the past 20 years allowing scientists to measure the abundance of metabolites and peptides in a sample . It measures the abundance of molecular ions coming from the ionization of metabolites or peptides/proteins present in a biological sample. Measuring metabolites or peptides/proteins defines two of the omics sciences that use this technology, metabolomics and proteomics, respectively. Most of the information shown here (section 5.1) is extracted from the book of Gary Siuzdak "The Expanding Role of Mass Spectrometry in Biotechnology" (*364*).

5.1.1. **Mass spectrometer parts**

Mass spectrometers are the instruments that perform the ionization, filtering and measurement of the ions present in a sample and are made of the following main parts: sample inlet, ionization source, mass analyzer and ion detector as shown in the schematic representation in Figure 20.

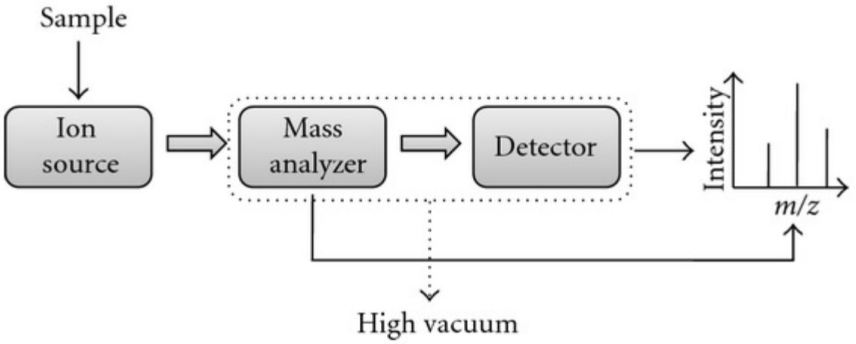


Figure 20. Mass spectrometer parts. Schematic representation of main parts in a mass spectrometer. From Banerjee and Mazumdar (*365*).

Sample inlet. The injection system that introduces the sample from a vial or chromatographic column into the ionization source.

Ionization source. The source of ionizing energy in which sample molecules in the gas phase are ionized by physico-chemical processes and converted to ions suitable to be analyzed in the mass spectrometer. The process involves energy transfer from the ionization source to the sample molecules, generating charged molecular ions from the original sample molecules. The ionization sources are classified in two main categories: soft ionization methods that leave the structure of the original molecule relatively unchanged, while hard ionization methods cause the fragmentation of the original molecule into smaller fragment ions.
Electrospray ionization (ESI) is a soft ionization method that is commonly used in liquid chromatography (LC) couplet to mass spectrometry proteomics and metabolomics. It uses a desolvation process (removal of solvent from the molecules of interest) by applying a high voltage current to produce charged molecular ions under very low pressure. Atmospheric pressure chemical ionization (APCI) and atmospheric pressure photon ionization (APPI) are other soft ionization sources less used to produce charged molecules when coupled with LC techniques.
Electron ionization (EI) is a hard ionization source technique highly reproducible and broadly used in gas chromatography (GC)/MS methods in metabolomics. Chemical ionization (CI) with different reaction gas, is another soft ionization source technique with a more narrow application spectrum in GC/MS metabolomics experiments, that has been proved of utility for tracing isotopes in labeling experiments (*366*).
Matrix-assisted laser desorption/ionization (MALDI) is the ionization technique used in MS imaging (MSI), which acquires MS spectra from bidimensional solid matrices (*367*). MALDI, desorption electrospray ionization (DESI) (*368*) and secondary ion mass spectrometry (SIMS)(*369*) are different ionization techniques with their own pros and cons that apply an ionization energy onto a bidimensional solid matrix instead of a liquid or gas solution, similar to MALDI.

Mass analyzer. Mass analyzers can be classified into two major groups: (1) trapping mass analyzers such as ion traps and orbi traps that resolve ions discontinuously, and (2) ion-beam mass spectrometers, such as ToF and quadrupole analyzers that resolve ions continuously.
There exist a variety of mass analyzers that combine static and dynamic electromagnetic fields to select and filter ions based on mass-to-charge ratio (m/z) resonance frequency, m/z stability or m/z flying

time, each of them with a different setup. Each instrument setup has its own acquisitions capabilities that can be tuned to obtain maximum instrument performance, only limited by the instrument setup. There are five key parameters that determine the limitations of the instrument setup: (1) **Sensitivity**, reported as limit of detection (LOD) and defined as the minimum amount of analyte that the analyzer detects; (2) **Resolution**, is the ability of separating two neighboring ion peaks. (3) **Scan rate**, defined as the frequency at which the instrument is recording ion abundances over a particular mass range, it goes from 1 to 100 Hz and depend on instrument architecture. (4) **Mass accuracy**, the ability of the mass analyzer to accurately provide m/z information, usually measured in parts per million (ppm) . (5) **Mass range**, the m/z range of the mass analyzer, a triple quad goes up to m/z 3000, or a time of flight (ToF) has theoretically unlimited m/z range.

Ion detector. After the ions are separated by the mass analyzer they arrive at the ion detector, which converts the incident ions to an electric current signal. For example, the electron multiplier transfers the kinetic energy of the incident ions to the surface that generates secondary electrons and electric signals. There are many types of detectors: electron multipliers, Faraday cup, photomultiplier conversion dinode, array detectors or charge detectors, each with its own limitations and advantages.

Instrumentation setup

A variety of instrument setups exist to meet the needs for mass accuracy, sensibility, resolution and mass range. For example, triple quadrupoles which have two quadrupoles with a collision cell in between them, offer the most sensitive instrument setup at the expense of a limited resolution and mass accuracy. However they are an excellent choice to quantify very small amounts of known small molecules in a targeted analysis. Hybrid instruments such as a quadrupole-Orbitrap, combine the efficiency of the quadrupole selecting the ions with the highly accurate and resolving power of Orbitraps, which makes them ideal for both peptide and small molecule detection/identification (*370*). The list of different instrument setups is still increasing with new hybrid instruments created to meet different needs and combining advantages of two or more different mass spectrometry instruments (*371*). Below in Table 3 extracted from Junot et al (2014) shows the main characteristics of most common MS setups.

Mass analyzer type		Resolving power (FWHM)	Mass accuracy (ppm)	Scan rate	Mass range	Dynamic range
Low resolution instruments						
QqQ		Up to 7500 (at *m/z* 508)	5-500 ppm	Up to ~5,000 Da/s (i.e. ~5 Hz when considering a 1000 *m/z* window)	Up to 3,000 *m/z*	10^5-10^6
IT/LIT		Up to ~10,000	50-500 ppm	Up to ~33,000 Da/s	Up to 4,000 *m/z*	10^4
Qq-LIT		Up to 9200 (at *m/z* 922)	50-500 ppm	Up to 20,000 Da/s	Up to 2,000 *m/z*	10^5-10^6
High resolution instruments						
TOF	TOF only	Up to ~20,000 (at ~*m/z* 1000)	< 1-2 ppm (internal calibration)	Up to 40 Hz	Up to 20,000 *m/z*	10^4-10^5
	Qq-TOF	Up to 60,000 (at *m/z* 1222)	< 1-2 ppm (internal calibration)	Up to 100 Hz	Up to 40,000 *m/z*	10^4-10^5
	IT-TOF	10,000 at *m/z* 1000	< 2 ppm (internal calibration)	10 Hz	Up to 5,000 *m/z*	10^3
Orbitrap	Orbitrap-Exactive	Up to 140,000 (at *m/z* 200)	< 1 ppm (internal calibration) < 3 ppm (external calibration)	Up to 12 Hz with a mass resolution of 17,500 (at *m/z* 200)	Up to 6,000 *m/z*	10^3-10^4
	LTQ-Orbitrap	Up to 240,000 (at *m/z* 400)	< 1 ppm (internal calibration) < 3 ppm (external calibration)	Up to ~8 Hz with a mass resolution of 15,000 (at *m/z* 400)	Up to 4,000 *m/z*	10^3-10^4
	Q-Orbitrap	Up to 140,000 (at *m/z* 200)	< 1 ppm (internal calibration) < 5 ppm (external calibration)	Up to 12 Hz with a mass resolution of 17,500 (at *m/z* 200)	Up to 4,000 *m/z*	10^3-10^4
FT-ICR	LTQ-FT 7T	>750,000 (at *m/z* 400)	< 1 ppm (internal calibration) < 1.2 ppm (external calibration)	1 Hz with a mass resolution of 100,000 (at *m/z* 400)	Up to 4,000 *m/z*	10^3-10^4
	Qq-FT 7T	> 1,000,000(at *m/z* 400)	< 1 ppm (internal calibration) < 1.5 ppm (external calibration)	1 Hz with a mass resolution of 250,000 (at *m/z* 400)	Up to 10,000 *m/z*	10^3-10^4

Table 3. Mass spectrometers in metabolomics. Overview of most commonly used mass spectrometers for metabolomics applications. From Junot et al (*372*).

5.1.2. **Tandem Mass Spectrometry (MS/MS)**

In mass spectrometry, the mass-to-charge ratio is a ratio resulting from dividing the compound mass by the number of charges present when ionized in the ionization source. This is a non-informative measure of molecular structure, hardly allowing the compound or peptide identification. In order to achieve structural information and be able to identify the small molecule or peptide, a higher degree of fragmentation is needed to allow the identification of the molecule of interest. And this is a fundamental step in the biomarker discovery and drug development process that mass spectrometry is allowing (*373*, *374*).

Tandem mass spectrometry, also known as MS/MS or MS2 involve three main steps: (1) selection of one or multiple ion/s, (2) fragmentation and (3) mass separation of the fragments. The fragmentation and separation of the resulting ions may take place in different compartments when the instrument have collision cells such as in QqQ, qToF or hybrid ion-traps, or in different scan times through ion accumulation in ion traps, Orbitraps or FT-ICR mass spectrometers.

The resulting fragments in a MS/MS experiment, allow the researcher to identify the original small molecule or peptide from where the fragments have originated, permitting its identification. In metabolomics, these fragments are highly specific to each metabolite with specific fragmentation patterns, but not unique (*375*). In proteomics, tandem mass spectrometry is used to fragment a peptide or protein in order to elucidate their amino acid sequence (*376*).

MS/MS acquisition modes

In MS/MS experiments there are three main different types of data acquisition, which can be identified depending on how ions are selected and MS/MS data acquired:

- Targeted mode: selection of known ions, acquisition of MS2 data from MS1 selected ions, obtain MS2 fragmentation information for few MS1 ions
- Data dependent mode: select ions based on a full scan MS1 criteria (intensity, mass accuracy, isotope pattern or neutral loss), acquisition of MS2 data for all MS1 ions selected
- Data independent mode: selection of MS1 ions based on m/z window, acquisition of MS2 data for all MS1 ions selected in MS1 increasing the complexity to link MS1 ion to MS2 fragments

5.1.3. **Targeted acquisition mode**

Targeted methods are characterized by having high sensitivity, wide dynamic range, reliable quantification and stability. With a targeted method you can measure the relative or absolute abundance of one or few analites at the same time, if a calibration curve is used for absolute quantification.
If the identification is not the objective, and one only wants to quantify a known analyte, two strategies allow the quantification of a preselected ion based on the full scan mode. Single ion monitoring (SIM) in a single quadrupole allows to monitor one precursor ion based on their previously characterized retention time (*377*), or monitor few precursor ions in a triple quadrupole using multiple ion monitoring (MIM) based on their retention times, in both cases with no fragmentation required.

However, MS/MS brings additional advantages since this technique can be applied to monitor multiple precursor ions and study their fragmentation patterns to achieve greater specificity. Here we found two different strategies very similar in their aim: multiple reaction monitoring (MRM) or selected reaction monitoring (SRM), typically performed with triple quadrupole, being the most used instrument setup in targeted metabolomics and targeted proteomics (*378*, *379*), or a slightly different approach which allows simultaneous scanning of an entire group of fragment ions in a high resolution instrument such as a quadrupole-Orbitrap (*380*).

5.1.4. **Data dependent acquisition mode**

Data dependent acquisition or DDA, is a technique that takes advantage of modern mass spectrometers which have fast duty cycles and acquisition times with tens of MS/MS scans per second. DDA is based on the ability of an instrument to filter and record some ions and a few milliseconds later perform a fragmentation (MS2) of the filtered ions in full scan MS1. The filtering process can be based on

different criteria: intensity of the MS1 ions, accurate-mass of the MS1 ions, isotope-pattern of the MS1 ions, or using pairs of low-high collision energy scans and monitor m/z differences between the low full scan and the high full scan. In untargeted metabolomics, DDA is more prone to technical errors, but can achieve greater metabolite assignment than the data independent acquisition strategy (*381*).

5.1.5. **Data independent acquisition mode**

Data independent acquisition or DIA, is another technique performed with modern mass spectrometers with fast duty cycles and acquisition times. In fact, the difference here is how the precursor ions are "filtered" to be fragmented. This approach provides fragment ion information for all precursor ions within a defined mass range. In DIA, high and low abundant ions are fragmented together, increasing the complexity of the fragmentation spectrum, and breaking the direct link between a specific precursor ion and its corresponding product ions. Turning the data analysis into a very complex and specific process but potentially with higher rewards (*382*, *383*).

5.1.6. **Separation techniques couplet to MS/MS**

In order to have more resolution, detect more ions and facilitate the extraction of ion abundance in a mass spectrometry experiment, a separation technique such as liquid chromatography (LC) or gas chromatography (GC) might be coupled to the mass spectrometer (*384*). These techniques allow the separation of the analytes present in the sample by using a stationary phase and a mobile phase. The stationary phase chemically and physically interacts and retains the analytes present in the sample, until the mobile phase gradually washes them away, eluting them from the chromatographic column. Each analyte has a specific retention time that depends on the strength of the interaction with the stationary phase and how the mobile phase interferes with this interaction. Separation techniques usually help reduce the matrix effects and ionization suppression, can separate isomers and facilitate a more accurate quantification of the analyzed metabolites.

5.2. Metabolomics

Metabolomics is the omic science that studies small molecules called metabolites. All the metabolites present in a cell, tissue, organism or sample are defined as the metabolome. We could see the metabolome as the end product and readout of gene transcription and protein activity. Transcripts and proteins are subject to epigenetic and post-translational modifications, making them more difficult to link to the phenotype. However, metabolomics is a powerful tool to analyze the abundance of known and unknown metabolites in a given sample and serve as a direct readout of biochemical activity in the cell or sample, and can be easily correlated to the phenotype. Therefore, it has become an important tool in basic, translational and clinical research, allowing researchers to elucidate some of the molecular mechanisms underlying a disease's onset, a drug treatment, the exposure to the environment, or the effect of a knockout gene in an animal or cell line experiment.

5.2.1. **Metabolomics workflow summary and metabolomics strategies**

A typical mass spectrometry-based metabolomics experiment workflow to measure some of the metabolites present in a given sample, follows in general the below steps:

1. Experimental design: the most important step to be able to generate interpretable and meaningful data to answer a biological question using metabolomics. The type of samples, experimental groups distribution, sampling method, data generation strategy or data analysis pipelines we are going to use for answering this question, are essential things to think about before starting a metabolomics experiment.
2. Sample collection: samples will be collected according to the type of metabolites we want to study. If the metabolites of interest need special collection methods (i.e. volatile metabolites), specific collection protocols exist to capture and preserve them such as solid-phase microextraction (SPME). In other samples, with specific metabolites of interest, we needed to add quenching or antioxidant substances such as ascorbic acid to prevent their degradation. In general, it is always good practice to immediately freeze the samples at -80 ºC after collecting them.
3. Sample preparation: often, samples need to be further processed before the metabolites are extracted. Sample homogenization, sample dehydration or sample lyophilization, or the addition of labeled standards, are common procedures before metabolite extraction.
4. Metabolite extraction: there are many different protocols and solvents to extract metabolites from a biological sample in a liquid-liquid extraction. Depending on which protocol and solvents one uses for the extraction, one will be able to detect and measure different metabolites. Metabolite extraction limits our ability to detect a family of metabolites or another, or a specific highly sensitive to degradation metabolite. To a certain degree, it will also determine the amount of metabolites, their nature and if the metabolites are more polar or more apolar. The process of metabolite extraction always includes a precipitation step where proteins, cell or tissue debris are precipitated to avoid any interference during chromatographic separation and data acquisition.
5. Metabolite/feature separation: because of the inherent high complexity of biological samples, it is often necessary to separate the metabolites present in the sample, by using gas (GC) or liquid chromatography (LC) techniques, before injecting the metabolites in the MS instrument. Separating the metabolites allows the MS instrument to increase its resolution, decrease the noise and be able to detect more non-overlapping points, allowing a better feature identification/annotation and metabolite quantification.
6. Data acquisition: once the metabolites/features/compounds are separated in the GC or LC column, metabolites enter the MS instrument and first are ionized (charged) by losing or gaining one or multiple charges in the ionization source. These ions are then filtered/scanned with a preset m/z range. Then, one or many known ions are selected (targeted), or all ions within a m/z range are selected (untargeted), using electromagnetic lenses that focus and filter these ions. At this point, selected ions can be fragmented in the collision cell (depending if we are acquiring MS1 or MS2 data). Then, the resulting ions are separated by their m/z ratio by different types of chemical and physical properties depending on the instrument setup used (time of flight, time being trapped in the orbital trap, selected ions in the quadrupole or ion

mobility). The final step involves the detection and counting of each ion that arrives at the detector, which converts chemical signals (ions) into electrical signals.

7. Data processing: after acquiring the raw data, spectral data points are processed by instrument vendor softwares or by open source softwares. For a targeted analysis the data processing step is simpler and aims to quantify/measure the abundance of the predefined selected compounds, once this is achieved we can perform the statistical analysis. For an untargeted analysis, the processing step is much more complex compared to the targeted counterpart. It usually involves peak detection, sample alignment, retention time alignment and feature filtering and correspondence, to allow a reliable compound identification/annotation and further statistical analysis. Note, that it is always recommended if present, correct the instrument variability along the run using the periodically injected quality controls (QCs) along the run.
8. Data analysis: once we have the matrix of metabolite features or targeted metabolites, we are ready to perform the statistical analysis. This typically involves descriptive statistics, batch detection and correction, and more complex analysis such as multivariate analysis or machine learning approaches.

I provide a more in depth explanation of the most relevant of these steps after defining the three main metabolomics strategies used in metabolomics in the next sections.

Targeted and Semi-targeted metabolomics

A targeted metabolomics experiment refers to a method in which a defined list of known metabolites is measured and is a hypothesis-driven strategy. It is usually created to test a biochemical or biological hypothesis that motivates researchers to investigate a specific metabolic pathway or a group of biologically interrelated metabolites. To create a targeted metabolomics method, an important analytical effort is necessary to optimize chromatographic separation and ionization responses of the targeted metabolites. Researchers use commercially available pure standards in order to optimize the chromatographic method and ionization parameters (*385*). These targeted approaches rely on highly sensitive and robust methods to measure tens of biologically relevant metabolites in a high throughput manner. It can measure low-abundant metabolites with good sensitivity and specificity, thanks to the optimized chromatography and the triple quadrupole technology used in this mass spectrometry approach (*386*). Additionally, targeted methods are quantitatively reliable and hence they are sometimes used to achieve absolute quantitation (*360*). The absolute quantification with this technology can be achieved for each targeted metabolite we want to measure, by constructing a calibration curve with the pure standard, and using a series of dilutions that cover the range of concentrations found in the samples measured.

A semi-targeted metabolomics method is another type of hypothesis-driven metabolomics experiment. We can see this approach as an extension of a targeted method, but used for the quantification of a much larger list of known metabolites (*387*). The optimization of this type of method is even more challenging than for a targeted method due to the use of hundreds of standard metabolites for the optimization and measurement. It usually involves different MS platforms and different extraction methods, as well as different chromatographic methods. It often covers high numbers of metabolites in the same family of metabolites such as lipids or carbohydrates as examples. This approach has the sensitivity of a targeted approach and can retain a comprehensive metabolic coverage without the drawbacks of untargeted metabolomics.

Untargeted metabolomics

An untargeted metabolomics method aims to measure as many unknown relevant metabolites as possible from biological samples, and is a discovery-driven or an hypothesis-generation approach. It aims to measure these biologically relevant molecular features that are differentially abundant in the samples after a rigorous data processing and statistical analysis (*360*). The advantages of an untargeted approach is that it allows the researcher to discover unknown or poorly characterized metabolites of biological or clinical significance. Hyphenated-MS platforms have become the setup of choice for most untargeted metabolomics experiments given their higher sensitivity. Metabolite quantification in an untargeted method needs a compromise between sample throughput, metabolome coverage and peak quality (*386*). Chromatography instruments are often coupled to high-resolution MS equipment that operates at a high scan speed since it is ideal to record the finest peak profiles possible during the chromatographic run. In untargeted metabolomics is common the use of orthogonal separation techniques coupled to HRMS, to enhance the metabolome coverage. For example, to cover polar and non-polar metabolites, HILIC-LC/MS and RP-LC/MS in addition to GC/MS techniques are used to expand the number and diversity of metabolites covered.

5.2.2. **Sample preparation and metabolite extraction**

An essential step for a metabolomics protocol is the extraction of the metabolites from the samples one wants to characterize, such as body fluids, cells, and fresh or fixed tissue. The objective in the extraction protocol is to obtain enough quantitative yields of metabolites from the sample to be able to detect them, and remove impurities, i.e. proteins or cell debris. The extraction protocol varies depending on the experimental design, the nature of the sample, the targeted molecules we want to detect with the analytical platform we want to use to measure them. In consequence, an optimal metabolite extraction method leads to a higher extraction efficiency and analytical sensitivity. However, an increased number of preparatory procedures and fractionations may reduce the analytic throughput (*388*). We can distinguish two main types of metabolite extraction methods: the liquid-liquid and solid phase extractions. Liquid-liquid extractions are more often used in targeted and untargeted metabolomics due to their superior throughput and reproducibility of results (*389*).

Extraction protocols aim to extract metabolites from the matrix where they are embedded/dissolved by exploiting a property that differs between the metabolites of interest and the rest of the matrix/metabolites (*390*). Therefore, sample preparation can be divided into:

- Enrichment for metabolites of interest: The most common extraction solvents include different combinations of water, methanol, acetonitrile, methyl tert-butyl ether and other organic solvents. It is important to note that sample extraction protocols target a subset of the metabolome present in a given sample. These extraction protocols are usually classified depending on whether the metabolites to be extracted are polar or nonpolar. But in general, we can say that an extraction protocol covering the entire metabolome does not exist (*391*).
- Removal of interfering impurities: Protein and cell debris precipitation is a necessary step to remove cell proteins/parts possibly interfering in the chromatography separation or metabolite ionization. This is achieved by mixing the sample extract for example with cold solvents such as

acetonitrile, ethanol or methanol in different pH conditions, that cause the denaturation and reduce the solubility of proteins and cell debris, for easier precipitation and removal.

5.2.3. **Metabolite separation**

The nature of the data detected depends on multiple factors. Metabolomics MS techniques can be divided depending on how the sample is injected in the mass spectrometer. Samples can be directly injected to mass spectrometers, a technique known as direct infusion (DI-MS). Or alternatively, a separation technique can be used prior to the injection to the MS, these techniques are known as hyphenated MS. Analytes that enter the mass spectrometer are previously separated by a gas chromatography (GC), liquid chromatography (LC) or capillary electrophoresis (CE) (*360*). Similarly to the diverse space of conditions for metabolite extraction, no single chromatographic method is suitable for all classes of metabolites and cannot cover the physicochemical properties needed to retain and separate the whole metabolome (*391*).
Another important technique to investigate metabolites present in biological samples such as tissues or organ slices, is imaging MS (IMS-MS). This technique allows ionization of metabolites from a solid surface, or solid-like matrix, providing information on the localization of the metabolite in the surface, by associating a specific mass spectra to a specific area in the surface. SIMS, DESI and MALDI are some of the MS techniques used to generate spectral data from tissue slices (*392*, *393*).

Liquid chromatography-MS metabolomics (LC/MS)

LC/MS metabolomics methods are able to cover the widest portion of the metabolome, from very small metabolites to large lipids or vitamins. This wide range comes from the variety of existent chromatographic columns, together with the sensitivity and accuracy of new instruments, that are typically used in targeted and untargeted metabolomics (*394*). Usually, LC/MS instrument setups used in metabolomics are composed by ultra-high pressure liquid-chromatography (UHPLC) coupled to triple quadrupoles (QqQ), hybrid high-resolution (HR) instruments such as quadrupole time of flight (qToF) or orbitraps, and where the standard ionization method is ESI. We have seen a recent increase in the development of LC columns that have greatly expanded their selectivity and improved separation efficiency, yet any chromatographic separation column is universal to separate the entire metabolome. For this reason, the combination of different chromatographic columns and mobile phases must be used to increase the coverage of a single LC/MS metabolomics experiment. There are two main types of chromatographic columns with their own diversity of stationary phases that allow chromatographic separation of the analytes, the reverse phase columns (RP) and the hydrophilic interactions chromatography columns (HILIC). And the use of different mobile phases in addition to extending or shortening the chromatographic time, can improve the separation efficiency and MS detection sensitivity (*390*).

Gas chromatography-MS metabolomics (GC/MS)

Volatile and nonvolatile low molecular weight polar and nonpolar metabolites are, or can be analyzed by GC/MS. Therefore, metabolome coverage by this technique is significantly lower than using its liquid chromatography counterpart. Often, GC/MS metabolomics methods rely on the application of a chemical derivatization step to analyze most polar and lower weight metabolites. Chemical

derivatization is applied to cover a broader range of the metabolome and allow more stability and volatility to the metabolites analyzed by GC/MS. One of the most common chemical derivatization approaches involve a two steps procedure, first the methoximation of the ketone groups, followed by the silylation of all protonated groups. There exists a great variety of column lengths, diameter, capillary types and stationary phase matrices that aim to enhance the coverage of the metabolome by this technique (*395*). Regarding the MS instrument couplet to the GC, different options exist such as QqQ, ToF, as well as HR instruments such as qToF or orbitraps. Electron Impact (EI) is the standard ionization method in these instruments that do not suffer from adduct formation or ion suppression that often occurs in LC/MS methods (*390*). Another interesting characteristic of GC/MS data is that the EI ionization is highly reproducible, facilitating the metabolite identification step.

5.2.4. **Metabolomics data acquisition**

Targeted metabolomics data is usually acquired using a triple-quadrupole or other instrument setups capable of selecting specific known m/z and performing or not MS/MS data. Nowadays, targeted metabolomics data analysis is a demanding and manual task supported by commercially available softwares distributed with the instrument. In contrast, untargeted metabolomics data is easier to generate with HR instruments but is highly complex and challenging to analysis process, and is usually analyzed by open-source software and demands a highly skilled data analyst to extract biological information (*389*). In a MS metabolomics experiment, the successively recorded ion histograms are stored in data files. Ionized molecules impact the detector during a short time (scan), and each histogram is constructed with the counts recorded of each ion of specific mass-to-charge ratio (m/z) and retention time (*396*). Generally, because targeted metabolomics monitors a relatively low number of precursor selected ions and its fragments, this type of data is less noisy than its untargeted counterpart. However, an untargeted approach monitors a higher number of precursor ions, and because of the intrinsic ionization dynamics, ion quantification and limitations of chromatographic separation, this type of data is more noisy and redundant depending on the experimental design and platform of choice. These inherent characteristics of untargeted metabolomics data must be considered carefully for the data analysis. Most common of these characteristics in an untargeted approach are:

- **Adduct formation**: during the desolvation process that occurs in electrospray ionization (ESI), especially in LC/MS but also with other soft ionization techniques, the formation of different adducts is common. Adduct formation can be favored but not controlled by varying the composition of the mobile phase. Adduct formation is molecule and origin dependent, the salt concentration in the mobile phase, the solvents or the sample matrix itself determine the adducts that will be recorded during the data acquisition (*397*). Dimers and trimers can appear if hydrogen bonds are highly expected within the molecule that is ionized in the ESI source (*397*). In addition, multicharged ions can be formed, increasing the complexity of the annotation process.
- **Isotopes masking**: all atoms have naturally occurring isotopes in more or less proportion, but natural isotope distribution of atoms like 13C or N15, 18O and 32S in high resolution instruments, must be also considered in the annotation process to avoid potentially masking of other ion signals (*389*).

- **Retention time shifts**: the chromatographic instrumentation is sensitive to temperature and atmospheric pressure changes. Therefore, if a very large number of samples are analyzed in the same run or in different batches, the same metabolite retention time for those samples can vary, and need to be corrected. Additionally, anomalies in the flow rate, mobile phase composition or age of the column might also shift the retention time.
- **Mass accuracy error**: the error tolerance depends on the instrument setup and its calibration, which will determine the window accuracy of the measured mass-to-charge ratio of the ions.
- **Noise to signal ratio**: there exists a considerable amount of chemical noise with different origins: column degradation, impurities from solvents and buffers or even from the sampled biological matrix. In addition, MS spectra contain random noise originating from the response of the detector to changes in the atmospheric pressure or temperature and its maintenance status.
- **Performance variation over time**: this refers to the global loss of response in the MS detector over time. When this happens, recorded counts may be lower than expected due to instrument usage and detector contamination and age (*389*).

5.2.5. **Metabolomics data processing in untargeted metabolomics**

There are three main steps in untargeted metabolomics data processing: baseline removal, peak detection and filtering, and retention time alignment. These steps are usually performed in instrument vendor or open source softwares after the conversion of vendor formats to an open data format known as mzML. The conversion to this format is usually the first step in untargeted metabolomics data analysis, and mzML is compatible with most open software to analyze metabolomics data. We obtain a matrix-like structure with three dimensions: m/z, retention time and intensity for each sample/file. And the difficulty of processing originates in the considerable variations in the chromatographic peak shape, peak with and m/z accuracy. The main processing steps mentioned consist in:

- **Peak detection**: peak peaking or peak detection consist in extracting as many mass spectrometry recorded signals as possible with a shape compatible with a gaussian-like shape. The algorithms used for this purpose often fit a model to an ideal gaussian shape of peak elution, and select peaks accordingly to similarity to the fitted model.
- **Noise filtering**: the algorithms used in this step aim to mathematically reduce the noise in the signal by subtracting this noise from the true signal in each peak.
- **Deconvolution**: deconvolution algorithms use mathematical operations involving two shaped functions. To distinguish two or more co-eluting compounds with similar peak shapes and retention time, these algorithms assume that fragments from the same molecule have the same retention time and profile across multiple samples and they are highly correlated because they are subject to the same biological and systematic variation (*398*, *399*). At the end, these algorithms allow us to estimate the relative area corresponding to each individual peak when multiple peaks are overlapping.
- **Sample and retention time alignment**: spectral alignment is a main step for processing metabolomics data. It aims to correct the retention time of nonlinear shifts in the peak and spectra that might occur when analyzing multiple samples and may considerably affect data quality. We can find two type of algorithms, the ones that align pairs of samples or multiple

samples spectra against a reference sample spectra or spectrum template, aligning the total ion chromatogram, or peak-based alignment or using an internal standard alignment; and other algorithms that use a reference-free alignment, they perform the alignment to the whole spectra or by splitting the spectra in smaller windows, and independently aligning each resulting segment (*400*). These algorithms apply binning, clustering and segmentation techniques, to align all the spectra in a sample.

- **Feature correspondence**: a feature is a measurable property of the object you are analyzing, a data variable. In untargeted metabolomics, features correspond to potential metabolites. Feature correspondence refers to the process of grouping sample ions with similar m/z across all samples with reproducible elution patterns. Retention time (RT) alignment and annotation accuracy error exist due to technical limitations. The objective of the feature correspondence step is to align peaks across samples using both m/z and RT feature values (*401*).

5.2.6. **Metabolomics data analysis**

During the processing steps, there is no distinction between adduct peaks, isotopes, in-source fragmentations or other impurities, and careful data analysis is needed (*402*). We can then use the processed data for statistical analysis, that consists in the selection of the data features detected that have significant variation across samples with a threshold of minimum abundances. Often, sample quantification needs to be adjusted due to instrument variability or processing batch, as well as values need to be transformed in a more comparable scale using normalization, scale transformation or data imputation.

- **Feature filtering:** because the ultimate goal of untargeted metabolomics is to identify metabolites, a minimum intensity threshold is needed to obtain reliable results. Low coefficient of variation (CV) variables are discarded for their potential lack of biological information.
    - Intensity filter: abundance threshold is one of the first filters that one can apply to reduce the complexity and enhance the metabolite identification. This filter aims to keep higher quality and higher ion intensities, allowing an acceptable ion intensity when identified by tandem mass spectrometry (MS/MS).
    - Quality control filtering: to control for the inherent technical variability in chromatographic and mass spectrometry instruments, metabolomic researchers inject periodically during the run, a sample composed by a pool of all the samples as a quality control (QC). This QC sample is injected across the run periodically and peak/features should have the same retention time and instrument response/intensity over the run (*389*).
- **Feature annotation**: feature annotation in untargeted metabolomics is a key step where different features may represent the same metabolite (*403*). This step is essential to reduce false positive rates and remove redundant information (i.e. isotopes, adducts and in-source fragments). Feature annotation algorithms aim to detect features corresponding to monoisotopic peaks of common adducts (i.e. H or NH4 adducts) (*404–406*). This algorithms contain three main steps:
    - Peak grouping and clustering: by peak-shape correlation or peak-abundance correlation
    - Cluster peak annotation
    - Neutral mass annotation

5.2.7. **Metabolomics feature identification**

There are five levels of identification confidence and different strategies for feature identification depending on the origin of the data and the strategy that is used to annotate the metabolite.

- **Compound identification**: to identify a metabolomic feature is the ultimate goal of untargeted metabolomics and constitute the main bottleneck in the analysis. However, feature identification is a challenging task which to a certain degree is hard to reproduce and standardize. We find five confidence levels of identification: 1) valid identification, this level assures unambiguous 3D structure elucidation including full stoichiometry. Needs the comparison of the identified compound to its pure reference standard in identical analytical conditions. 2) incomplete identification, this level differs from level 1 because the pure standard was not analyzed under identical analytical conditions, but can confirm 2D structure. 3) putative identification, this level has some certainty on the formula and part of its structure, but can not supply a metabolite identity. 4) molecular formula annotation, here a single molecular formula corresponds to multiple candidate structures with adduct and charge information. 5) Unknown compounds, MS/MS spectra cannot be annotated
- **GC-EI-MS identification**: in most GC-EI-MS experiments, the ionization energy is set to 70eV, the fragmentation patterns produced at this ionization energy are highly reproducible facilitating compound identification (*395*, *407*). Retention indices (RI) are commonly used to allow a more confident identification, injecting a collection of alkanes that help to normalize based on their retention time. RIs in combination with EI spectral matching, significantly improve metabolite identification (*408*). The spectral matching is commonly used against the national institute for standards and technology database (NIST), the human metabolome database (HMDB), MassBank and other databases containing EI spectra.
- **LC/MS identification**: the confidence level of identification in LC/MS based identification is commonly 2 or 3, because the 1st confidence level of identification is out of reach for most of the laboratories. In general, experimental MS/MS spectra is investigated and matched against a reference MS/MS database aiming to find a definitive match. But some considerations need to be taken into account before matching the MS/MS spectrum such as:
  a)Acquisition instrument type: QqQ, qToF, ion trap and orbitrap produces different fragmentation patterns. b)Mass accuracy can affect both precursors and fragments spectra. c)Ionization mode and energy can also affect spectra quality. d) Adduct of the precursor ion can also influence the acquired spectrum.
- **Automatic feature identification**: given the enhanced capabilities of DDA systems to obtain hundreds of MS/MS spectra in a single run, the process of metabolite identification needs to be automated. The algorithms need to first filter candidates from the experimental spectra matching a reference spectra, based on mass accuracy, ionization parameters and similarity to the reference spectra. The similarity between experimental and reference spectra is calculated with the relative intensities of experimental and reference spectra and other weighting algorithms (*409*). However, sometimes there is no reference spectrum available and an in-silico prediction of fragmentation spectra may be used.
- **In-silico MS/MS**: given a compound structure a few computational algorithms have been developed to predict their MS/MS spectra. First, the chemical structure needs to be encoded in

a computer readable format, then the in-silico MS/MS can be calculated, although prone to miscalculations. For calculating this MS/MS spectra, four methods can be distinguished: 1) quantum chemistry: uses chemistry principles; 2) machine learning: requires training sets; 3) heuristic algorithms: limited to specific compound families; 4) reaction chemistry: based on reactions in metabolic pathways.

Once we have metabolite identities or putative identifications, we can perform different types of enrichment analysis to detect altered pathways under the biological challenge of the treatment group versus the control, or between the different experimental groups. As well as perform standard, univariate and multivariate statistics to characterize the biological sample.

5.2.8. **Metabolomics isotope tracing and flux analysis**

Stable isotopes like carbon C^{13} and nitrogen N^{15} are less abundant but present in nature. They have the same number of protons and electrons, but different in the number of neutrons, yet conserving the same chemical formula, structure and properties. But the presence of these extra neutrons confers these stable isotopes with an "extra" weight that can be used to differentiate them from unlabeled ones, in a similar way as to radioactive labeling experiments, but not hazardous (*403*). There are a few naturally abundant atoms such as carbon, hidrogen, nitrogen, oxygen or sulfur with relevant abundances.

The main advantage of tracing stable isotope labeled molecules (isotopologues) for metabolomics experiments relies on the ability to study the dynamics of metabolic fluxes when a labeled molecule, for example labeled glucose, is used as primary source of energy in a cell culture experiment. These experiments allow one to trace their metabolic fate through the metabolism, by uniformly or positionally labeling a molecule such as glucose, and following the labeled atoms in a targeted or untargeted manner through the metabolism or specific pathway (Figure 21) (*410*, *411*).

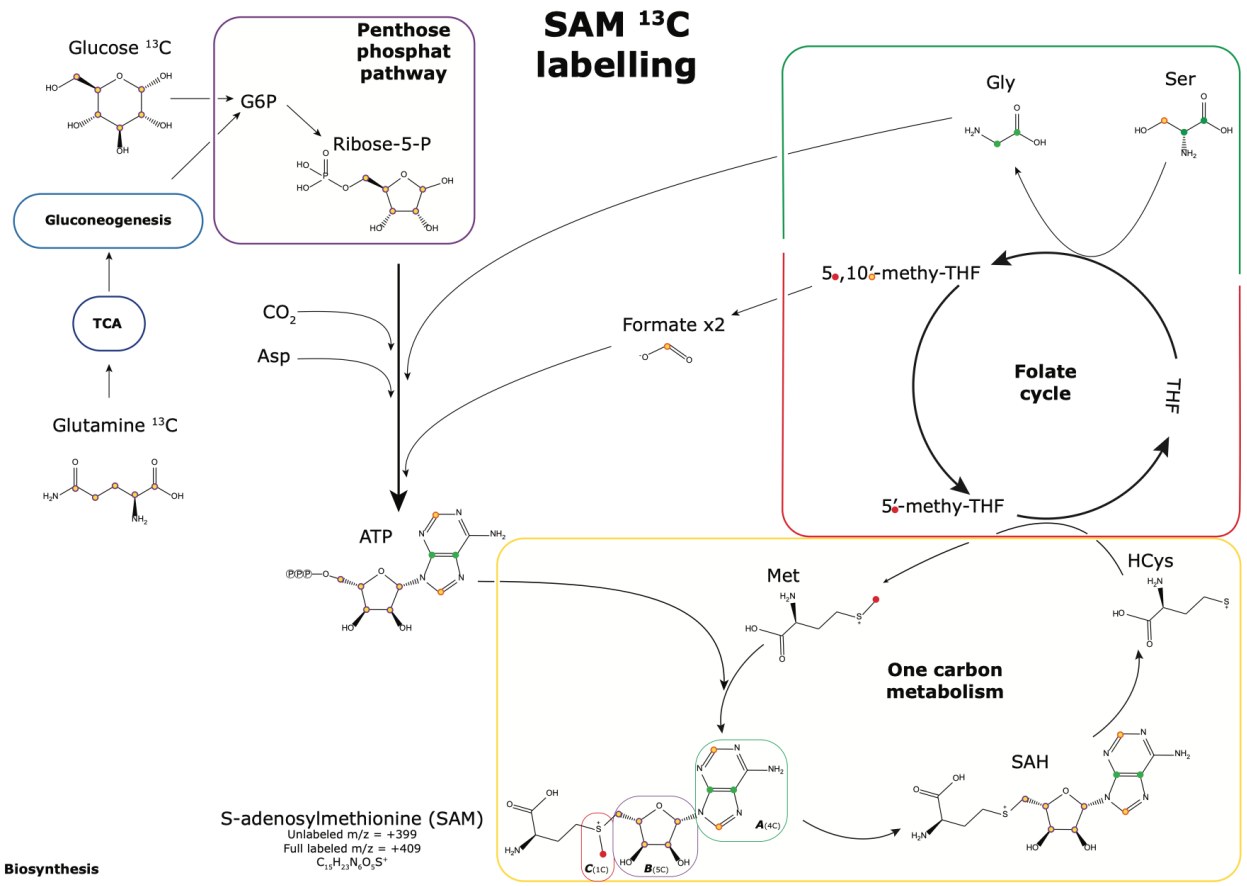


Figure 21. S-adenosylmethionine (SAM) labeling. Stable isotope labeling of glucose or glutamine and the connection to the synthesis of S-adenosylmethionine. From publication in Chapter 2.

To choose the optimal stable isotope labeled tracing experiment, one needs to take into account two extra considerations. First, the tracer metabolite of choice (i.e. glucose), depends on the question one wants to answer (experimental design), and the knowledge of metabolic pathways involved in its metabolization, one will choose one or another tracer metabolite. Second, the kinetics of the studied system will determine that the isotopomers reach a detectable concentration allowing them to measure and estimate its flux (*412*).

Qualitative stable isotope labeling flux estimations generate information about the origin and rates of production of downstream metabolites with steady pattern. While, a quantitative estimation wants to quantitatively assess the absolute flux in a specific pathway or metabolic network. To achieve the latter, one needs to mathematically model the metabolic reaction and fluxes that govern the metabolization of the trace metabolite up to the labeled secondary metabolites, using genome-scale metabolic reconstructions and quantitatively estimating fluxes (*413*, *414*).

5.3. Proteomics

Proteomics is the omic science that studies proteins. All the proteins present in a cell, tissue, organism or sample are defined as the proteome. We could see the proteome as the effector of most functions in a cell, tissue or organism. Proteins are incredibly diverse and are formed by the linear combination of around 20 amino acids, where their linear order and location in the 3D structure of the peptide, leads to

a very diverse range of proteins. Their conformation, isoforms and post translational modifications (PTMs) affect their functionality and stability, and are the object of study in proteomics research. Similarly to metabolomics, the proteomics field has grown in terms of throughput and improvement measurement thanks to mass spectrometry advances (*415*), although very different strategies may be used to study them.

5.3.1. **Instrumentation in proteomics**

Electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) are the two ionization techniques used in mass-spectrometry based proteomics. The first step in a proteomics workflow aims to reduce the sample complexity by separating, fractionating or enriching the proteins and peptides present in a complex biological sample or protein digest, previous to MS analysis (*415*). Due to the high complexity of protein mixtures and protein digests from a biological sample, proteomics uses two-dimensional or even three-dimensional chromatographic separation of the peptides present in the sample, by combining RPLC, HILIC, ion exchange or affinity chromatography separation (*415*).

5.3.2. **Proteomics strategies**

There are two mainly proteomics strategies: top-down and bottom-up

- Top-down proteomics aims to analyze intact proteins, it has less ambiguities than bottom-up strategies and can distinguish between isoforms. It is normally applied to highly purified protein samples such as a single protein or simple protein mixtures (*416*).
- Bottom-up proteomics or shot-gun proteomics, aims to profile highly-complex samples for large scale analyses. Proteins are enzymatically digested into peptides (i.e. digested with trypsin) prior to MS analysis. The main drawbacks of this strategy are the difficulty of elucidation of a protein from its peptide sequences, limited coverage of a protein and loss of labile PTMs (*415*).

5.3.3. **Quantitative proteomics**

MS peptide abundance or relative abundance has been used to compare abundances across samples in a adimensional manner using protein ratios. But, recent advances in labeling strategies, MS and bioinformatics can now allow the absolute quantification of proteins in a sample, the protein copy number per cell *(417)*. Below, the main strategies for the quantitation of labeled sample proteins:

- Stable-isotope labeling with amino acid in cell culture (SILAC): proteins are labeled by culturing cells in media containing 15N salts or 13C-labeled amino acids, easy to apply and compatible with purification procedures.
- Isotope coded affinity tag (ITAG): proteins are labeled with chemical probes that consist of three elements: a reactive group, an isotopically labeled linker and a tag (i.e. biotin) for the affinity isolation of labeled peptides.
- Isobaric tags for relative and absolute quantification (iTRAQ): proteins are labeled at specific sites, the reagents may contain affinity tags for its isolation after protein digestion.

Labeled-free strategies are also a valid option despite suffering from the above mentioned drawbacks in mass spectrometry, regarding linearity and instrument variability. Absolute protein quantification can be achieved with targeted proteomics methods using SRM, MRM or PRM methods. However, a limited coverage is achieved because of the need of manually curating each peptide for each target protein (*418*).

5.3.4. **Protein identification**

The main bottleneck of protein identification comes from inferring protein entities from peptide spectras generated in the LC-MS/MS experiment (*376*). In high throughput experiments protein identification can be performed in-silico with two strategies: 1) searching existing spectral databases or directly searching the experimental spectra and reconstructing the peptide sequence; 2) protein identification is solved by calculating the probability of identification when the combination of m/z precursor and its fragment ions are matched to known peptide sequences in large protein databases.

5.4. Next Generation Sequencing (NGS) technologies

Sequencing technologies have improved dramatically since Frederic Sanger in 1977 developed the first sequencing method (*419*) or since in 2001 the first publication of a draft human genome sequence was published (*420*). Next generation sequencing (NGS) technologies have revolutionized life-science research, decreasing the costs per sequenced base dramatically in the last decade. Nowadays, NGS instruments are able to sequence at the same time, and only in a few days, tens to hundreds of whole human genomes, a remarkable increase in throughput and decrease in the price per Gb sequenced. In general, we can classify sequencing technologies in three types, the first generation of sequencing technologies are Sanger sequencing-based technologies, the second generation of sequencing technologies are the ones that currently are the most used worldwide like Illumina sequencing, and a third generation of sequencing technologies are the ones with promising simplified workflows and increased genomic gathered information, such as Nanopore or PacBio technologies, each of them with their advantages and limitations. The sequencing of a DNA or a RNA molecule consists in reading and enumerating the linear sequence of the nucleotide bases present in the sequenced molecule to identify genes and other genomic features. To achieve that, there are common steps shared between the second and third generation technologies that are summarized below:

1. **Sample preparation**: the DNA or RNA is extracted from the cell, the nucleus, a tissue, the whole organism or sample of interest. This genetic material is purified and fragmented into smaller pieces or not, typically ranging from a few hundred to several thousand base pairs in length, depending on the sequencing platform being used. The fragmentation can be performed enzymatically or mechanically using sonication, depending on the research application. The size of the DNA or RNA fragments is critical to construct a good quality sequencing library and ensure optimal sequencing results, the length will depend very much on the sequencing platform and configuration of choice.

2. **Library preparation**: the isolated and fragmented genomic material is then processed to construct a library of fragments for sequencing. In this and previous step relays the huge variety of genomic characteristics/features that a researcher can measure with NGS technology. This step aims to prepare the genomic material to allow its sequencing in the platform of choice, while at the same time unveiling a specific genomic characteristic. Library preparation usually involves several steps but not limited to end repair, A tailing, adapter ligation, and amplification. End repair and A tailing are performed to repair the ends of the purified DNA or RNA fragments to ensure that they can be ligated to the adapters correctly. Adapters are short pieces of DNA that contain sequences necessary for the sequencing and identification of fragments and samples. After adapter ligation, PCR is used to amplify the library, which generates multiple copies of each fragment for sequencing. Some technologies or applications avoid PCR amplification to reduce bias and polymerase errors.

3. **Sequencing**: the sequencing step consists in reading the genomic information present in the fragment/library construct one base at a time, enumerating the linear sequence of bases. When the sequencing library is loaded into the sequencing machine, each base is readed following different strategies depending on the platform used. There are several sequencing technologies available, such as Illumina, Ion Torrent, PacBio, or Oxford Nanopore, among others. Each technology has its own strengths and limitations in terms of read length, accuracy, and throughput. For example, Illumina sequencing is the most widely used and offers high accuracy and throughput, but has shorter read lengths that make the analysis more complex, while PacBio sequencing offers long reads with high accuracy but at very expensive costs, or Oxford Nanopore technology offers long reads but with higher error rates in the base calling process.

4. **Data processing and analysis**: the first step in the data processing is quality control, which involves assessing the quality of the raw data and filtering out low-quality reads. The reads are then trimmed to remove any adapter sequences or low-quality bases, depending on sequencing technology used. Then, reads are aligned to a reference genome or assembled de novo to generate new contigs and scaffolds, to finally base calling and analyzing the different variants, which may identify single nucleotide polymorphisms (SNPs), insertions, deletions, or structural variants among other features that depend on the sample and library preparation. As well as the quantification of the genetic features present in the library (i.e. transcript counts or DNA gene copies)

5. **Data interpretation**: the final step in any sequencing experiment is data interpretation, which involves analyzing the results of the sequencing experiment in the context of the biological question being addressed. For example, the data may be used to identify disease-causing mutations, study gene expression patterns after a treatment, or characterize microbial communities in the human gut. The interpretation of the data requires specialized expertise and computational resources, such as bioinformatics tools and databases.

Some sequencing techniques can be applied in single cells. Bulk techniques have been developed firstly, and characterize any given biological sample containing nucleic acids in a population-level mixt of cells. For example, if we study the transcriptome of an homogenized liver biopsy piece, we are measuring an average transcriptome of all cells present in that piece, similar to a population study. While in contrast,

if all cells of the liver piece are disaggregated and separated in single cells, a more grained and precise transcriptome is drawn for the entire pieces, highlighting differences between cell populations. Unfortunately, not every sequencing platform, protocol or application is suited for single-cell sequencing, in addition to a more complex data analysis process which needs to deal with zero inflated results.

5.4.1. Sequencing technologies basic principles

Sanger sequencing

Sanger sequencing is the first generation of sequencing technologies. Is the first method for DNA sequencing developed in the 1970s by Fred Sanger and his colleagues (*419*). The method involves using chain-terminating dideoxynucleotides (ddNTPs) to terminate DNA synthesis, which generates a set of fragments of varying lengths that can be separated by size using gel electrophoresis (Figure 22). This sequencing technique has been automated to gain throughput capabilities and reduce its costs. In general involves five main steps:

- **DNA template isolation**: extraction of the temple DNA from the sample of interest, or its amplification through a PCR reaction.
- **Primer design**: primers must flank the targeted region of interest.
- **Sequencing reaction**: involves adding the template DNA, designed primers, DNA polymerase, deoxynucleotides (dNTPs), and a small amount of one of four ddNTPs (ddATP, ddCTP, ddGTP, or ddTTP) together into four separate reaction tubes. The ddNTPs are labeled with different fluorescent dyes, which allows the fragments to be detected and distinguished from one another during electrophoresis.
- **Electrophoresis**: The four sequencing reactions are then run on a gel electrophoresis apparatus, which separates the fragments based on size. The fragments are detected using a laser or UV light, and the signals are captured by a detector.
- **Data analysis**: data generated by the sequencing experiment is analyzed by vendors and open softwares that can align the sequences to a reference genome, identify variants such as SNPs, and provide information on the quality of the sequencing reads.

Automated versions of the Sanger sequencing method have been developed, increasing substantially the throughput and number of samples sequenced at the same time.

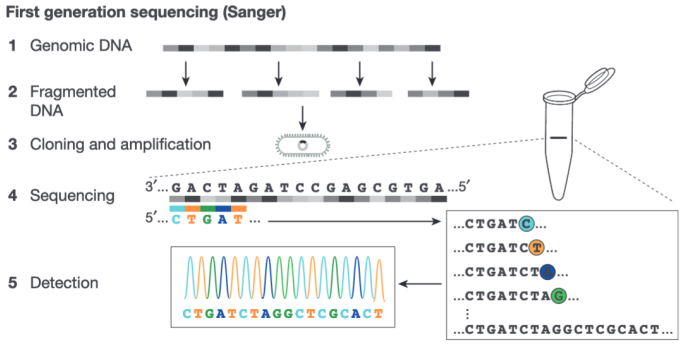


Figure 22. Sanger sequencing overview. Schematic representation of the Sanger sequencing method and main steps in the protocol. From Shendure et al. (*421*)

Illumina sequencing

Illumina and other equivalent sequencing technologies (i.e. Ion proton, 454) are the second generation of sequencing technologies, or also known as next generation sequencing (NGS). Illumina sequencing technology is the most widely used second generation sequencing technology that uses a sequencing-by-synthesis (SBS) chemistry, to generate millions to billions of short DNA sequences in parallel. It differs from the Sanger method because it tracks the addition of nucleotides as the DNA chain is copied, instead of the Sanger chain-termination method. It generates short fragments, and compensates its error rate by increasing dramatically the number of short DNA fragments sequenced, finally obtaining an accurate consensus sequence. In general this technology involves the following steps:

- **Template**: extraction of the temple RNA or DNA from the samples of interest, or its amplification through a PCR reaction is needed for library preparation.
- **Library preparation**: aims to process the RNA or DNA fragments for the construction of the sequencing library compatible with the sequencing platform and configuration of choice. It usually involves fragmenting the extracted DNA or RNA, converting the RNA to complementary cDNA, repairing the end and A-ligation to allow the correct ligation of the adapters, together with the addition of the sample barcode (indexes). Most protocols perform a library amplification step by PCR to increase the signal and number of clusters formed from the same molecule. The final step in the library preparation is a quality control and library quantification to ensure optimal loading and cluster generation.
- **Cluster generation**: aim to create clusters of DNA fragments on the surface of the flow cell, each cluster originating from a single molecule as shown in Figure 23. This is done by attaching the sequencing libraries to a coated layer of oligonucleotides in the surface of the flow cell that are complementary to one of the two adapter sequences present in the constructed library molecules. A single molecule attached to the complementary sequence of the adapter in the surface of the flow cell, is then amplified by bridge amplification. This variation of a PCR consists in copying the molecule followed by the bending and formation of a "bridge" towards the second complementary oligo in the surface. This process generates clusters of identical DNA fragments that are spatially separated on the flow cell surface. In pair-end runs (PE), when read 1 has been sequenced, another bridge amplification step is performed in the sequencing instrument to allow the second read, read 2 to be sequenced. Over- or under- loading the flow cell results in overlapping or too distant clusters, losing output capacity because of poor quality or not-enough clusters present in the surface, respectively.
- **Sequencing-by-synthesis**: the process involves the addition of a mixture with the four fluorescently labeled nucleotides (A, C, G, and T) to the flow cell, and by competition, only the complementary nucleotide will be attached by the sequencing polymerase. Because each nucleotide is labeled with a different color of fluorophore, the base that is being incorporated into the growing DNA strand can be identified, by imaging the flow cell and detecting each cluster color. The nucleotides are added in a stepwise manner, one base at a time, and are detected using a high-resolution camera. When a nucleotide

is incorporated into the DNA strand, the fluorescent signal is recorded and the nucleotide is cleaved from the DNA strand, allowing the next nucleotide to be added in the following cycle.

- **Imaging and base calling**: after each cycle of nucleotide incorporation is performed, the flow cell is imaged using the high resolution camera to record the fluorescent signal from each cluster. Fluorescent signals are then analyzed by a software to determine the identity of each base that was incorporated into the DNA strand. This process is called base calling, and the resulting data is called a sequencing read. The reads are typically 150-300 base pairs in length, depending on the sequencing platform and configuration being used.
- **Data analysis**: the first step is quality control, which involves assessing the quality of the raw data, where adapters are trimmed, duplicate rates are estimated or low-quality reads are filtered out. In a PE run, read 1 and read 2 then are linked to allow an optimal alignment to a reference genome or easier the novo assembly. Finally, variants are called and analyzed, which may involve identifying single nucleotide polymorphisms (SNPs), insertions, deletions, or structural variants, as well as detecting the levels of a particular transcript or mapping where a transcription factor binds, to mention a few application examples.

This technology offers a wide variety of library preparations and sequencing short-reads options, from stranded protocols to know the strand where the transcript/read was originated from, to DNA methylation or microRNA analysis protocols, to the analysis of any custom library preparation constructed with compatible sequencing chemistry.

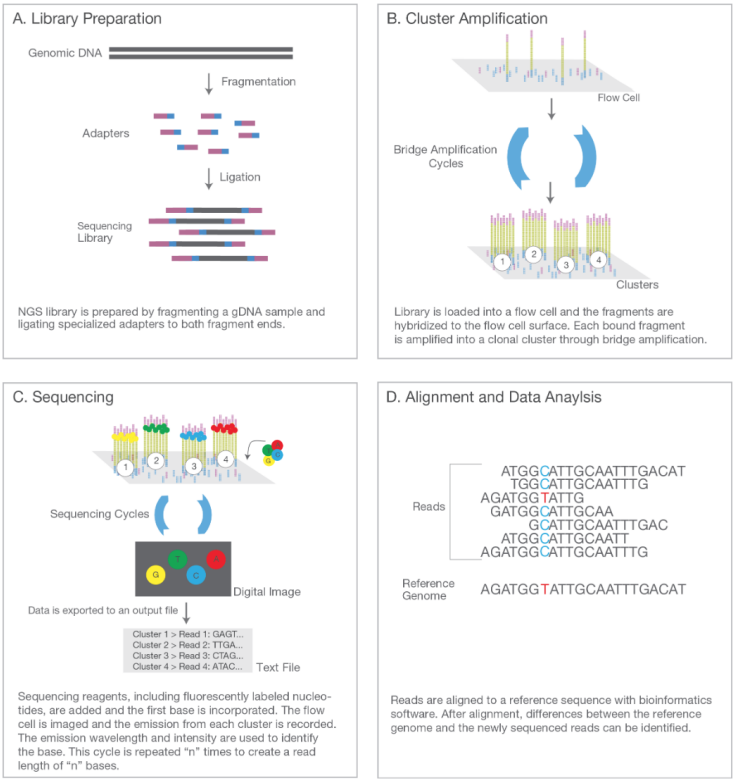


Figure 23. Illumina Sequencing chemistry overview. Overview of illumina chemistry from library preparation to data analysis. Library preparation where the genomic material is used to construct a sequencing library (A). Cluster generation where clusters of copies form the same library are attached to the sequencing flow cell (B). Sequencing of the clusters enhances the signal (C). Alignment to the genome and data analysis (D) .From Illumina.com

SMRT and Nanopore sequencing

These two technologies belong to the third generation of sequencing technologies. Both of those technologies form a group of sequencing methods that generate long reads, sometimes even spanning entire DNA molecules or RNA transcripts, which can help overcome some of the limitations of short-length sequencing read technologies. In addition, both have unique features allowing them to fill a specialized niche in the sequencing technologies ecosystem. SMRT sequencing produces the highest accurate long reads, and Nanopore allows the detection of few modified bases and has miniaturized/made portable the sequencing device to a USB-like size, and both generate sequencing data from one single molecule at a time.

- **Nanopore sequencing**:
  Oxford Nanopore technology relies on nanopores embedded in an electro-resistant membrane, which are connected to a sensor chip that records the electric current that flows through each nanopore in the flow cell. When a molecule passes through the nanopore, the electric current flowing through the nanopore is disrupted, and base calling algorithms call each base based on this electric change. Each nucleic acid base modifies the current differently, meaning that modified bases can also be identified. The nucleic acid molecules are being directly sequenced with no required PCR step, preserving modified bases. Another feature of this technology is that it sequence one molecule in each nanopore, allowing single molecule sequencing. In general, this technology involves the following steps:
    - **Template**: DNA/RNA is extracted of the highest molecular weight and purity possible, and optionally step of fragmentation to smaller sizes can be performed.
    - **Library preparation**: first, the ends of the molecule are uniformly repaired enzymatically by making the 5' end a phosphate, and adding an adenine overhang to the 3' end. Next, the adapters or alternatively, cleavage and addition of transposase adapters, are added to each molecule. Adapters are attached to a motor protein that controls the DNA/RNA strand movement through the nanopore during sequencing. Together with the motor protein, a hydrophobic tether is used to localize the template molecule near the nanopore for optimal sequencing performance.
    - **Sequencing**: the library is directed to the nanopore and the molecule starts to pass through the nanopore, at that moment, the base-calling algorithm calculates the nucleotide base present in a given moment, based on the changes in the electric current through the nanopore in real time. Converting these electric changes in nucleotide identities, the base caller constructs the linear nucleotide enumeration of the molecule being sequenced.

- **Data analysis**: the data analysis will involve in general removing low quality reads, trimming the adapters and mapping to the reference genome or assembling a new one. But in general, long reads will easily detect any variants present, speeding up the biological insights obtained from this data.

Oxford Nanopore Technologies (ONT) offers different library and sequencing options depending on the nucleic acid length, research application and available complementary lab instrumentation. One of the main advantages of ONT is the portability that offers one of their miniaturized devices, scientists have already brought them to Antarctica, near the sample collection places in the middle of the African savanna, or even into the international space station (Figure 24).

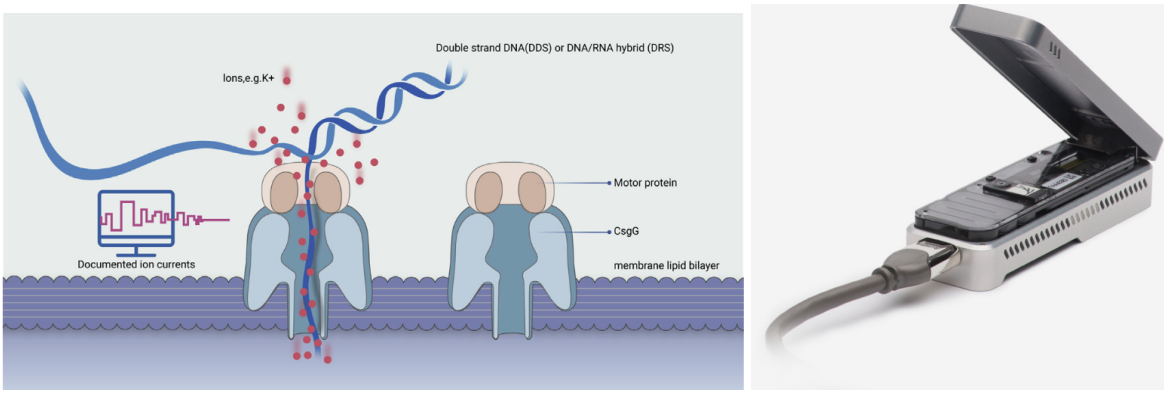


Figure 24. Nanopore sequencing technology. Sequencing nanopore representation (a). From Xie et al (*422*). minION devices have a USB size (b). From nanoporetech.com

- **Pacific Bioscience sequencing (SMRT)**:
  Pacific Bioscience sequencing technology differs from Nanopore technology, because it relies on the construct of a circular DNA (or cDNA) template that is incorporated and immobilized in very small wells in the sequencing flow cell (*423*, *424*). Single molecule real time (SMRT) sequencing technology sequences a single molecule in each microwell and captures the light emitted in the well, when a new base is incorporated and the fluorophore is liberated (Figure 25). The high fidelity in copying the template molecule is achieved with multiple rounds of sequencing the same circular molecule. SMRT sequencing generally involves:
  - **Template**: DNA/RNA extraction of the highest molecular weight and purity possible, and optionally fragmented into smaller sizes.
  - **Library preparation**: SMRT library preparation consists in capping the nucleic acid fragments on both sides and ligating hairpin adapters, which creates a circular template where the polymerase can circulate and incorporate nucleotides to generate the sequencing read. SMRT library prep also allows samples to be indexed and multiplexed to increase throughput.
  - **Sequencing**: the sequencing flow cell contains millions of very small wells called zero-mode waveguides (ZMWs) which allocates a single library construct per well. When the polymerase incorporates each complementary nucleotide to the template, light is emitted and recorded when the fluorophore is released, and the base calling algorithm identifies the base and constructs the read sequence.

An advantage of this circular template is that a single molecule can be sequenced multiple times generating a highly accurate long read. Additionally, this technology can also measure some epigenetic modifications in the nucleic acids based on the kinetics of base incorporation, when a modified base is present.

- **Data analysis**: similar to Nanopore, PacBio reads data analysis will involve in general removing low quality reads (if any), trimming the adapters and mapping to the reference genome or assembling a new one. Long high fidelity (HiFi) reads will easily detect variants present, with more accuracy than its nanopore counterpart due to the higher accuracy and lower error rate in base calling.

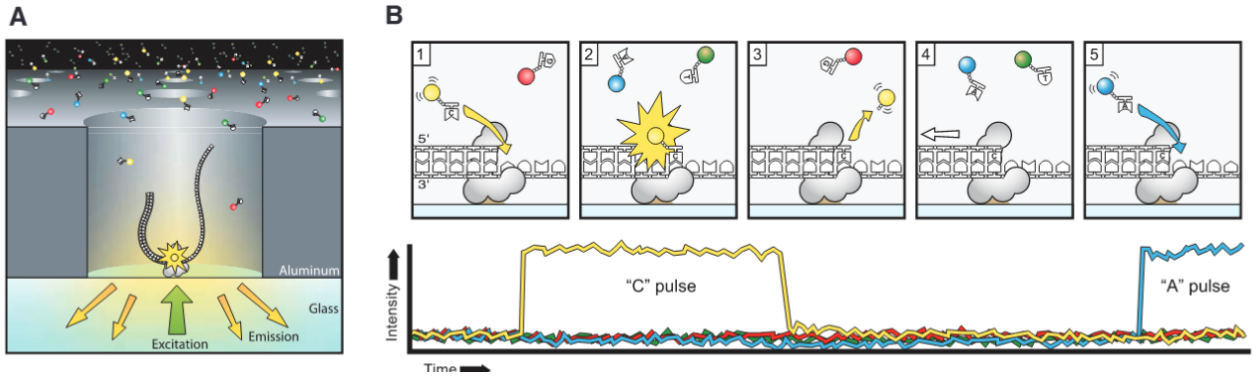


Figure 25. Single molecule real time (SMRT) sequencing. SMRT sequencing in a zero-mode waveguides (ZMWs) well where the sequencing happens (A). Schematic representation of how the sequencing signal and chemistry works (B). From Eid et al (*424*).

The three technologies presented above are not the only ones that exist today, but yes, they represent the most widely used and with the most exciting potential applications of all of them. From massively sequencing hundreds to thousands of human genomes in a few days, to obtaining highly accurate very long reads, or to bringing the sequencing device into the field to directly analyze a freshly collected sample.

There are many NGS applications in the genomics era to study a variety of genetic and epigenetic features in the genome, transcriptome and epigenome. Just to enumerate a few of the most common protocols applied to study any given sample: whole genome sequencing (WGS) to study the whole genome and detect variants; transcriptome sequencing (RNA-seq) to study how the genome is transcribed; chromatin-immunoprecipitation sequencing (ChIP-seq) to study how proteins interact with DNA or map a specific DNA/RNA modification; DNA methylation (WGBS) to study the DNA CpG dinucleotides methylation status in the genome; exome sequencing to study the different coding exons in an organism and how they are arranged; 16S metagenomics sequencing to study the microbial community present in a given ecosystem based on a marker gene; high-throughput chromosome conformation capture (Hi-C) to measure the frequency of two different DNA regions being physically associated in the 3D space; the assay for transposase-accessible chromatin (ATAC-seq) to measures the chromatin accessibility across the genome. By no means this list is exhaustive, and it only wants to give you a short overview of the variety of protocols and features that can be studied with NGS technologies.

During my PhD research, I have used 3 of these NGS applications. RNA-seq, WGBS and 16S metagenomics to study how the transcriptome, DNA methylation and microbial communities change in the different groups of my mouse model. Below I describe how each technology works and the main steps in the generation and interpretation of the data.

5.4.2. **Transcriptomics (RNA-seq)**

Sequencing the transcriptome (RNA) of a given sample has several advantages over sequencing the whole genome (DNA): it can give insights of how the genes are expressed, estimating the expression levels of individual genes, allowing the comparison between conditions and identifying genes that are up- or down-regulated in a specific condition (*425*); it can identify transcript isoforms resulting from alternative splicing events, allowing the identification of functionality diversity in the expressed genes (*426*, *427*); it can also identify novel transcripts that have not been previously annotated in the genome, helping researchers to discover new genes or identify previously unknown regulatory elements (*425*); it can detect fusion events between different genes, which can be indicative of oncogenic events in cancer cells (*428*): or can be used for pathway and enrichment analysis of groups of genes that are functionally related, gaining insights into the biological pathways and processes that are affected by different conditions or treatments. A technical description of the main steps followed to generate transcriptomics data are described below:

1. **RNA isolation**: the total RNA is extracted from the cells or tissues of interest using various methods, such as phenol-chloroform extraction, magnetic bead-based purification or commercially available extraction kits.

2. **RNA quality assessment**: the extracted RNA is quantified and quality is assessed using various methods, such as gel electrophoresis, UV spectrophotometry, or a bioanalyzer, to ensure that the RNA is of sufficient quality for downstream applications and calculate the RNA integrative number (RIN), a measure of the grade of degradation of the extracted RNA.

3. **rRNA removal**: due to the high proportion of ribosomal RNA in the transcriptome (many copies across the whole genome), it is recommended to removed this kind of RNA that can negatively impact the sequencing by diminishing the proportion of on-targeted total RNA (non-rRNA).

4. **RNA fragmentation**: the extracted RNA is fragmented into smaller pieces using enzymatic or mechanical methods.

5. **cDNA synthesis**: the fragmented RNA is converted into cDNA using reverse transcriptase enzymes and random primers or oligo(dT) primers.

6. **cDNA library preparation**: the cDNA fragments are then ligated with adapters that contain unique molecular identifiers (UMIs), sequencing primers and sample indexes for multiplexing. UMIs allow the removal of PCR duplicates and accurate quantification of transcript abundance.

7. **Sequencing**: the cDNA libraries are then sequenced using NGS sequencers. During sequencing, the adapters and UMIs are used to identify individual cDNA fragments and assign them to specific transcripts and samples.

8. **Data analysis**: the raw sequencing data is preprocessed to remove low-quality reads and adapter sequences. Then, the reads are aligned to a reference genome or transcriptome using software tools such as *STAR*, *HISAT*2, or *Bowtie*2 (*429–431*). Finally, transcript expression levels are quantified using software tools such as *HTseq*, *Kallisto*, or *Salmon* (*432–434*).

9. **Differential expression analysis**: the quantified expression levels or counts are then used to perform the differential expression analysis with softwares such as *DESeq*2 or *edgeR* (*435*, *436*). Which in general fit a negative binomial function and estimates the dispersion based on size factors of the count matrix for each gene, and apply a statistical test to calculate p-values and fold changes.

10. **Functional analysis**: differentially expressed genes can be then interpreted using functional analysis tools, such as gene ontology (GO) analysis or pathway analysis, to identify the biological processes and pathways that are affected by the differentially expressed genes in each comparison between conditions.

Overall, RNA sequencing is a powerful technology that allows a transcriptome-wide detection and quantification of the genes that are being transcribed from DNA. It allows the measurement of low and high abundant transcripts with a high dynamic range. It has the ability to detect novel transcripts and alternative splicing events, which can provide insight into the regulation of gene expression and the functional diversity of transcripts. RNA-seq data can be used to perform functional enrichment analysis, which allows the identification of biological processes, molecular functions, and cellular components that are altered in different conditions (i.e. control versus treatment). And this data can be integrated with other omics data such as metabolomics, proteomics, metagenomics or epigenomics to provide a more comprehensive view of the biological system.

5.4.3. **DNA methylation (WGBS)**

Whole genome bisulfite sequencing (WGBS) is the gold standard technique to investigate the methylation status of all CpG dinucleotides present in the genome at single nucleotide resolution. DNA methylation is an important epigenetic feature that controls gene expression and cellular function, therefore its characterisation and study allow important biological functions to be interpreted. Usually, hypermethylation of the DNA has been linked to the repression of genes (*252*), because transcription factors are sensitive to these modification (*437*), it has also been linked to less open chromatin states (*438*), and is a mark of oxidative stress (*439*), aging (*440*), cancer development (*441*) or neuronal decline (*442*). In contrast, hypomethylation is a key feature during early embryonic development (*443*), and in some cancers (*444*). Therefore, the study of the DNA methylation status can give a view complementary to RNA-seq, ChIP-seq or ATAC-seq data among other applications. A more technical description of the steps followed to generate WGBS data is described below:

1. **DNA isolation**: high-quality DNA is extracted from the cells or tissues of interest using various methods, such as phenol-chloroform extraction, magnetic bead-based purification or commercially available extraction kits.

2. **Bisulfite conversion**: in order to differentiate methylated from unmethylated cytosines, a bisulfite chemical reaction is performed to convert all unmethylated cytosines (C) to uracil, and leave methylated cytosines unchanged. This will allow the identification of which C are methylated and which not when compared to the reference genome. Unmethylated cytosines converted to uracil in the bisulfite conversion, will be transformed to a thymine during the library preparation and sequencing steps.

3. **Library Preparation**: bisulfite-treated DNA is then fragmented to an appropriate size for the sequencing platform. DNA fragments are then end-repaired, ligated with adapters, and subjected to PCR amplification. Then, library quantification and quality control is performed to ensure optimal sequencing performance.

4. **Sequencing**: in WGBS, libraries are sequenced in standard sequencing instruments and cover almost all CpGs in the entire genome. Due to the higher abundances of the T nucleotide after bisulfite conversion, special care should be taken when multiplexing samples to ensure optimal sequencing performance.

5. **Methylation Calling**: the read depth, number of times a nucleotide in the genome has been sequenced, is used to determine the methylation status of each cytosine, by comparing the bisulfite-converted DNA sequence to the reference genome. The methylation status can be quantified as the ratio between the number of reads containing a methylated cytosine divided to the total number of reads covering that specific cytosine.

6. **Data Analysis**: methylation data can be analyzed using a variety of bioinformatics tools to identify differentially methylated regions (DMRs) between samples, to correlate DNA methylation patterns with gene expression, or to identify genomic features that are enriched for DNA methylation, such as promoters or enhancers.

7. **Functional analysis**: one can use the annotated genes near the identified DMRs regions or CpGs studied to perform a functional analysis, such as gene ontology (GO) analysis or pathway analysis. This allows the identification of the biological processes and pathways that are affected by the hyper- or hypo-methylated regions or CpGs between conditions. However, not all DMRs will be influencing the nearby gene, some will have long-range interactions functioning as distal enhancers for other genomic regions far away.

Overall, WGBS can provide valuable biological insight into the epigenetic regulation of gene expression and the functional consequences of DNA methylation. The identification of DMRs define regions differentially methylated between samples and conditions that can be potential biomarkers for disease diagnosis and treatment. But individual CpGs can also be used as a biomarker in specific diseases. When integrated with other omics, for example with RNA-seq data, DNA methylation can identify genes regulated by the methylation status of their promoters, enhancers or their gene bodies. DNA

methylation information can serve to study inheritance patterns from parents to offspring, or to investigate epigenetic mechanisms of development, or study its role in diseases.

5.4.4. **16S Metagenomics (amplicon sequencing)**

In order to understand and analyze the microbiome and thanks to the advancement of next generation sequencing, mass spectrometry technologies and new computational tools we can now interrogate the microbiome dynamics of a given sample at an unprecedented resolution (*445*, *446*). To study the microbial community dynamics, we use animal models or take a small sample of the human population, and sequence or culture that microbiome and study its composition, phylogeny and functionality. However, microbiome experiments are often difficult to reproduce in different labs because of the huge variability in the microbiome composition between two individuals (*447*), and sometimes even between two biological replicates of a controlled experiment (*448*). For that reason an appropriate experimental design for the question of interest, strict control of the experimental conditions in animal models, proper sample collection and storage, suitable definition of control and inclusion criteria, appropriate sample extraction, modeling possible technical variation, and suitable sequencing technology are needed to successfully extract biological meaningful information from microbiome studies (*449*). However, many researchers are unaware of the compositional nature of any type of metagenomics data such as marker gene 16S ribosomal gene, shotgun metagenomics or metatranscriptomics. Compositional data contains information about the relationship between the entities that form a dataset, rather than the absolute values (*450*). We could see a high throughput experiment as a space with a fixed-size number of slots, and where these slots are filled with a random sampling of the relative abundance from our ecosystem of study (*451*). This implies that compositional data are inflated in zeros and need to be dealt with specific tools for compositional data, rather than standard approaches that are less optimal in compositional data.

Microbial communities are present everywhere and have gained attention in the biomedical research community due to its discovered importance in many diseases and healthy homeostasis. There are three main techniques to study the microbiomes or metagenomes present in a given sample using (i.e. the mammalian gut microbiota): 16S amplicon sequencing, which aims to detect the presence and identify the microbial species/phylotypes present in a sample, based on a marker gene; shotgun metagenomics, which aims to detect all genes (including the marker gene), to identify and measure the genetic functionality encoded in the whole microbial community; or metatranscriptomics, which consist in studying the genes that are being expressed in a given microbial community. Basically, shotgun metagenomics aims to sequence as many whole microbial genomes as possible, while 16S amplicon sequencing aims to identify and quantify which phylotypes are present and their abundances, or metatranscriptomics, aims to sequence the transcriptome to measure the functionality being expressed in a given moment. A more technical description of the steps followed to generate 16S amplicon sequencing marker gene data is described below:

1. **Template:** DNA/RNA is isolated from the sample or environment of interest with specialized protocols ensuring good quality of the starting material.

2. **PCR amplification**: the 16S rRNA gene is amplified using PCR (polymerase chain reaction) with universal primers flanking one or more hypervariable regions of the conserved 16S rDNA

gene, used for microbial classification. The PCR primers amplify a region of about 550bp and can be used to construct the sequencing library after its purification.

3. **Library preparation**: PCR products are then ligated with adapters that allow them to be sequenced on the chosen sequencing platform configuration. Library quantification and quality control is performed to ensure optimal sequencing performance.

4. **Sequencing**: libraries are then loaded onto the sequencing platform and the DNA fragments are sequenced, usually in a pair end (PE) mode. Because microbes have small genomes, it is recommended to sequence long fragments in PE mode, to facilitate its de novo assembly. The combination of short reads and long reads technologies brings substantial advantages for the novo assembly.

5. **Data analysis**: raw sequencing data is then analyzed using bioinformatics tools to identify the microbial taxa present in the sample and their relative abundance. This involves comparing the sequences to a reference database of known 16S rRNA sequences to assign taxonomy and generate a taxonomic or phylotype profile of the sample.

Overall, 16S amplicon sequencing can provide the diversity, structure and dynamics of a microbial community in a sample. It has several advantages over shotgun metagenomics such as providing taxonomic resolution in a high throughput in a cost-effective manner, enabling the identification of microbial taxa and phylogeny, describing the microbial community structure and specific diversity distribution between conditions. By identifying a taxon one can infer the functional potential encoded within their genome. It also allows the calculation of several diversity metrics within and between samples such as richness, alpha diversity or beta diversity.
I describe the main steps of the microbiome data analysis and deepen the indexes explanation.

Microbiome data analysis

There are many tools and strategies to analyze microbiome data depending on the type of data and research question. But, there are common steps in both, the preprocessing steps and statistical analysis that can be generalized. The first unavoidable steps in all NGS experiments is the demultiplexing of the samples using the index information and distinguishing reads for each sample, the removal of primer and adaptor sequences, application of a quality filter to remove poor quality reads and merge pair-end reads if needed. Then, we need to map the reads to a reference database to assign a taxon (16s), a gene (metagenomics) or a transcript (metatranscriptomics), to identify the operative taxonomic unit (OTU), the genes present, or the transcript expressed in a specific microbiome sample, respectively. Bioinformatic tools used to perform this first steps in the analysis of microbiome data are: dada2 (*452*), phyloseq (*453*), Qiime2 (*454*), bowtie2 (*429*), SortMeRNA (*455*), MetaPhlAn4 (*456*), HUMAnN2 (*457*) or MEGAN (*458*) are some examples of tools used to pre-process and assign phylogeny or taxon to microbiome data and obtain a "count" matrix .

From the moment we obtain these "count" matrices we are dealing with compositional data, where many zeros are present and need to be taken into account during the downstream analysis calculating distances, correlations and dissimilarities between samples. A required step in any statistical analysis is

the normalization of the data, where data is usually adjusted to a normal distribution, but as compositional data is inherently "normalized", read count normalization is unnecessary. Instead, a log-ratio transformation of the data makes the data symmetric and linearly related, and places the data in a log-ratio coordinate space (*450*). Then the next step in a traditional analisis, after rarefaction or count normalization, is the calculation of the dissimilarity or distances matrix. Typical dissimilarity and distance metrics are UniFrac, Bray-Curtis or Jensen-Shannon divergence, where none of them take into account the compositional nature of the data. The calculation of correlations using these metrics are not accounting for the compositionality of the data, putting a severe problem to use correlations and covariances for ordination, clustering, network analysis or differential (relative) abundance determination (*451*). Benchmarking of the traditional tools used for microbiome analysis have shown that false discovery rate depends on different levels of sparsity (*459*), and the false discovery rate can be up to 20x higher than expected (*460*).

Despite the challenge of working with compositional data in metagenomics research, there are some metrics such as richness, evenness, Shannon index, alpha-diversity or beta-diversity that can help us to compare samples and characterise the microbiome of an ecosystem (as described in section 3.1.4). Species richness measures the number of species detected in an environment or ecological community and does not take into account the abundance of the species. Species evenness, also called Pielou index, is a measure of the diversity together with the richness, where the evenness is the count of each individual for each species (*461*). Species richness is subject to the sampling effort, then different rarefaction techniques commonly used in microbiome analysis, highlight different aspects of biodiversity changes (*462*). The Shannon index is a measure of diversity and takes into account the number of species present in a sample (richness) and their relative abundance (evenness) (*463*). Alpha-diversity is a measure of the mean species diversity at a local scale in a given ecosystem and time point. Beta-diversity is the difference in alpha-diversity between two locations or individuals. And the gamma-diversity is the total diversity in a landscape or ecosystem (*464*). All those metrics can help one to characterize microbiome data and add valuable biological and ecological information.

5.5. Multi-omics statistical analysis and integration

Multi-omics data generation and data integration aims to study the relationship between the different variables present in each molecular layer. High throughput molecular layers data (omics) generated in the same set of samples, brings an unprecedented opportunity to learn the regulatory principles that unites each of the molecular layers in the holobiont. Because each omics dataset is often generated by different technologies, with different statistical properties, different possible bias, and nature of the signal, careful needs to be taken when studying the omics relationships. To help studying these complex relationships, many computational tools and statistical frameworks have been developed (*465*). They help processing, normalizing and transforming the data to have optimal formats for data analysis and integration. The statistical frameworks used, aim to mathematically decompose the system variance and simplify/reduce the axis of variation for easier interpretation. The algorithms used in these frameworks take many mathematical strategies to integrate multi-omics data (Figure 26), all with strengths and limitations. They can be based on correlations or similarity metrics, using Bayesian approaches or integrating the different omics layers using a multivariate statistical approach, among others methods

(*465*). One essential characteristic needed for almost all multi-omics integration is that the omics measurements need to be measured in the same set of samples (*466*), partially overlapping samples (*467*, *468*), or can also integrate different omics layers by measuring the same set of variables in different sets of samples (*469*). Measuring different variables in the genome, proteome, metabolome or microbiome from the same individual samples, allow an holistic integration of all the measured molecular layers and study their relationship in an integrative manner.

Two other important aspects to choose the integration method are: if one knows a priori the omics dataset distribution (parametric) or if it is not know (non-parametric); and whether one wants to integrate the different dataset in a supervised (unlabeled data) or supervised manner (labeled data) (*466*). There are different valid tools to integrate different omics data in the python and matlab programming language (*470*, *471*), but most of the integration tools developed lately are implemented in the *R* programming language (*468*, *469*, *472–474*), which is open and freely available to use and is accompanied with many complementary *Bioconductor* R packages.

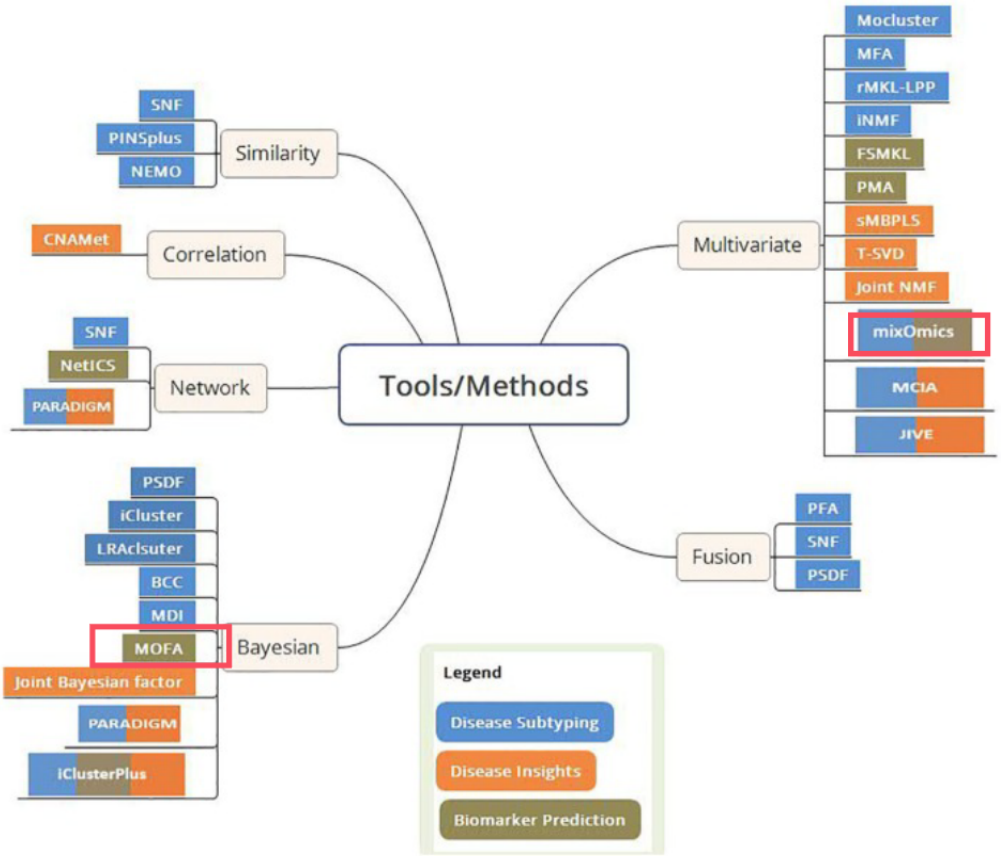


Figure 26: Overview of multi-omics data integration tools. The tools/methods are grouped and colored based on their approach and applications. *FSMKL*: feature selection multiple kernel learning; *JIVE*, joint and individual variation explained; *MCIA*, multiple co-inertia analysis; *MDI*, multiple dataset integration; *MFA*, multiple factor analysis; *MOFA*, multi-omics factor analysis; *NEMO*, neighborhood based multi-omics clustering; *PFA*,

pattern fusion analysis; *PMA*, penalized multivariate analysis; *sMBPLS*, sparse multi-block partial least squares; *SNF*, similarity network fusion; *NMF*, nonnegative matrix factorization; *BCC*, Bayesian consensus clustering; *PSDF*, patient-specific data fusion; *mixOmics*, multivariate supervised and unsupervised integration. Adapted from Subramanian et al. (*465*).

During the development of this PhD thesis I have used several *R* packages and tools from the *Bioconductor (475)* and *CRAN* projects to process, normalyze, transform, explore and characterize the different omics data sets generated on the germ-free mice model in light of the holobiont. I have used descriptive univariate statistics together with different visualization strategies to describe a single variable behavior in a given data set. But, I have also used multivariate descriptive statistics with more complex visualization strategies to simultaneously observe and study the behaviors of multiple variables at the same time in a single molecular layer. However, the ultimate goal of this PhD is to integrate the different measured omics layers and study their relationships, to better understand the mouse holobiont. The integration of different molecular omics layers has allowed me to better characterize the holobiont and observe the interconnections between the different molecular layers.
To accomplish the integration at the multi-omics level, I have mainly used two *R* packages: the first is called multi-omics factor analysis (*MOFA*/*MOFA*+) that use an unsupervised strategy to integration the different omics (*467*, *468*); and the second *R* package is called *mixOmics* for the supervised omics integration and exploration (*469*). The sections below aim to describe the principal features of these two approaches to integrate multi-omics experiments in a single statistical framework.

5.5.1. **Multi-Omics Factor Analysis (MOFA, unsupervised)**

The *MOFA*2 package has been developed to characterize and integrate single-cell multi-omics experiments (*468*). However, many examples already published in high impact journals have demonstrated its utility to integrate bulk multi-omics data sets as well (*476–478*). Before constructing the model, each omics data set needs to be transformed to be properly normalized, filtered to keep only highly variable features, and undesired source of variability removed (i.e. batch effect), for optimal performance of the algorithm. Here are the principal characteristics of the *MOFA* statistical framework:

- Allows the simultaneous integration of omics-data measured in the same or partially overlapping set of samples in an unsupervised statistical fashion.
- Accommodate the integration of time-course experiments and composed of multiple experimental groups.
- Permits to uncover the underlying factors driving the observed biological variability, by learning mathematically inferred latent factors that capture the global variability of the system.
- Is an unsupervised statistical framework, meaning that no prior data knowledge (i.e. experimental groups) is used by the model to learn the latent factors.
- Is a Bayesian-base probabilistic model, which uses the measured omics to calculate the posterior probability given a prior probability distribution, using a linear model.
- Missing values are allowed in the input data and can be or not imputed.
- The interpretation of the latent factors is analogous to the interpretation of the principal components in a principal component analysis (PCA), but for a multi-omics experiment.

- Input data need to be properly normalized, filtered and batch removed, to avoid the model capturing undesired source of variability, and variance stabilized.
- Because it is a linear model, it is subject to missing strong nonlinear relationships within and across the different omics layers.

On the whole, *MOFA* provides a flexible statistical framework to integrate different omics modalities in an unsupervised fashion (Figure 27). Allowing to capture the principal axes of data variability through learning the inferred latent factors by decomposing and assigning part of the variability explained in each factor, to each data modality feature. It might happen that a factor is only specific to a data modality or that the factor is shared between some of the data modalities. In summary, MOFA is a powerful Bayesian linear model which allows the detection of most relevant technical and/or biological variability in the system, and provides an easy to implement framework for results interpretation.

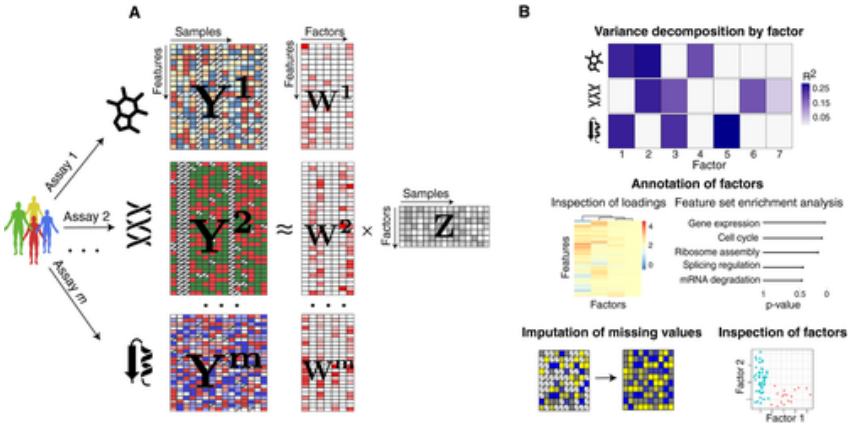


Figure 27. Multi-Omics Factor Analysis. MOFA model overview and schematic downstream analyses. From Argelaguet et al (*467*).

5.5.2. **mixOmics (supervised)**

The supervised approach, using the *mixOmics R* package, is focused on data exploration, dimensionality reduction and data visualization (*469*). It provides various statistical methods to integrate several omics datasets and find the relationship between them. The main methods on the package are based on partial least square regression (PLS) and canonical correlation analysis (CCA), which aims to find the relationship between 2 or more matrices and maximize their covariance in a more or less supervised manner; and principal component analysis (PCA) as its unsupervised approach. However, one needs to first properly transform and normalize the data, and prefilter zero variance variables similarly to *MOFA*. Here are the principal characteristics of *mixOmics*:

- Aims at exploring, mining, integrating and visualizing multi-omics high throughput data sets.
- It implements supervised and unsupervised approaches that allows the horizontal or vertical integration of different omics layers, measured in the same N-samples (N-integration) or measuring the same P-predictors in different samples (P-integration), respectively.
- Aims to classify or discriminate sample groups to identify the most relevant features in a multi-omics experiment, maximizing the features covariances between two or more matrices.

- Its main multi-omics N-integrative approach for variables measured on the same set of samples is called DIABLO, and is based on a multiblock sparse PLS discriminant analysis framework.
- Both, P- and N-integrations aim to identify biological relevant and robust molecular signatures.
- Allows the tunning of multiple parameters for optimal feature and component selection.
- Provides several functions to assess the performance of the different methods.

Overall, *mixOmics* offers several tools for the supervised and unsupervised integration of multi-omics data measurements in the same (N-integration) or different (P-integration) set of samples (Figure 28). Latent components are constructed with linear combinations of the selected features for each omic layer measured to reduce complexity and enhance biological interpretation. In general, *mixOmics* has implemented several unsupervised and supervised approaches to explore and characterize the samples.

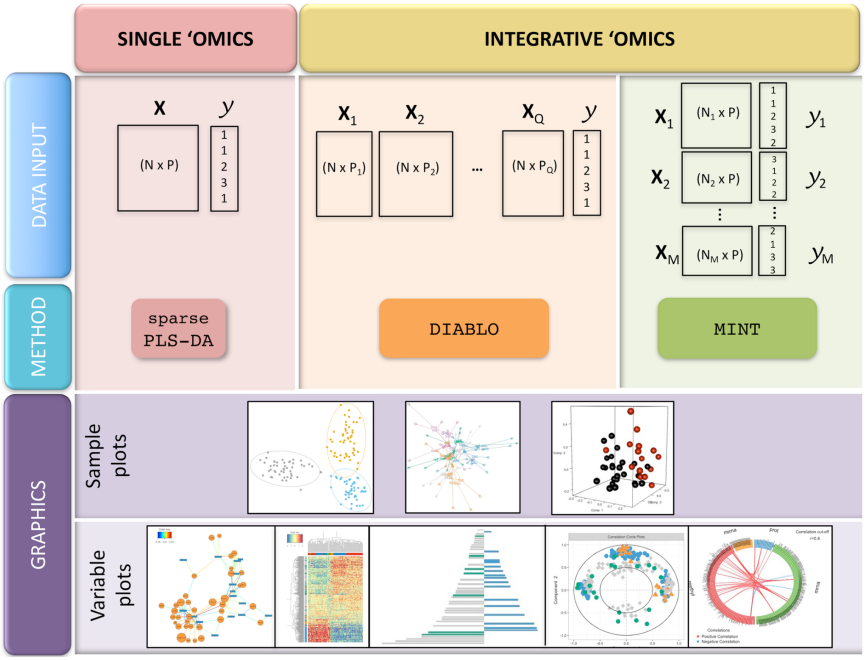


Figure 28. mixOmics overview. Multivariate methods for single and integrative omics supervised and unsupervised analyses. From Rohart et al. (*469*).

6. Mice in microbiome research models

Animals ranging from fruit flies, or worms to chimps, have historically been used to answer fundamental biological questions, and often to attempt to respond to human physiological, genetic or biochemical questions. Among them, mice are one of the most popular animal models used to model human biology. Thanks to its genetic and physiological similarity to humans, their small size, relatively quick life cycle and being less expensive than higher mammals, mice are often used in translational biomedical research projects.

However, extreme care should be taken when translating results found in a mice model to human biology, or when comparing results from different inbred mice strains or handled in different facilities (*479*). For example, there are several differences between the mouse and human gastro-intestinal tract such as the stomach pH or colon segmentation, and only the intestinal layout is shared between them. Although a similar gut microbiota composition is observed in both intestinal tracts, the relative abundance of taxons differs significantly between mice and humans, and some time between mouse strains (Figure 29). Remarkably, over 70% of the annotated functions encoded in both microbiomes (mouse and human) are overlapping, indicating that most of the microbiome functions are shared between mice and humans (*479*). Another cautionary tale is the different genetic background among mouse strains. These differences can significantly affect the composition of the gut microbiota, the response to diet or the response to environmental changes. Biological sex of each mouse should also be considered when modeling human diseases, because sex-specific microbial signatures have been reported in respect to hormonal homeostasis, immune regulation, circadian gene expression among other biological functions (*102*, *170*, *480*).

Although the translation of results from mice studies to humans should be regarded carefully, mouse models are still one of the best options to model human diseases. The section below is not an exhaustive list of microbiome models, but yes the most used ones.

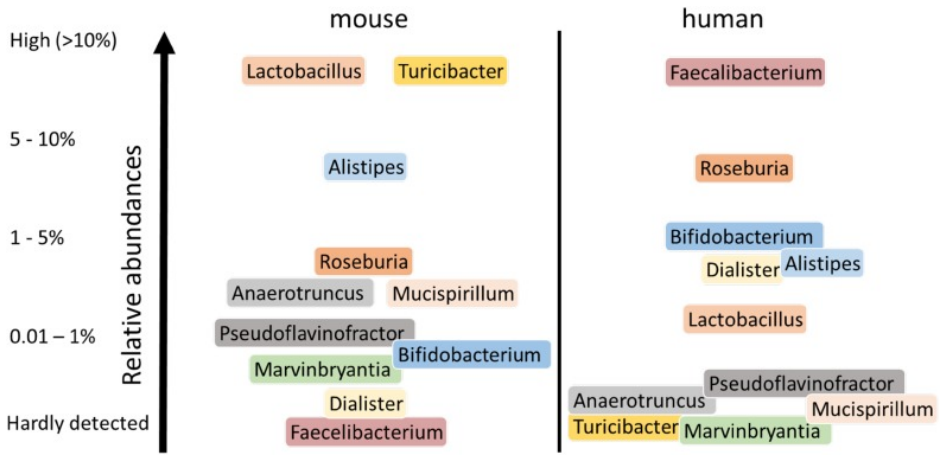


Figure 29. Major different human and murine intestinal genera. Only genera are shown that showed consistent differences in relative abundance between humans and mice. From *Hugenholtz and de Vos* (*479*)

6.1.1. **High fat diet (HFD)**

These models aim to reproduce and model obesity, type 2 diabetes, metabolic or immune diseases, administering a high fat/high sugar diet to the mice. Mice under this diet have significantly reduced diversity and an increased abundance of certain bacterial species that are associated with inflammation and metabolic dysfunction (*481–483*). In addition, interventions that modulate the microbiome, such as prebiotics, probiotics, and fecal microbiota transplantation (FMT), have been shown to have beneficial effects on HFD-induced obesity and related metabolic disorders (*484–486*).

6.1.2. **Germ-free (GF)**

Together with colonization strategies, supplementation with bioactive compounds or treatment with fecal microbiota transplantations, germ-free (mice raised with no microbiota) models are very useful to investigate the complex relationship between the host and its microbiota at many levels. They provide a platform where mechanistic experiments can be tested. They can help to understand how the immune system is trained, how microbes (or its absence) help to gather energy from nutrients, how the gut-brain axis works, and a long list of possibilities exist (*487–489*). But they allow more causative and mechanistic insight when combined with colonization of single or small bacterial communities, or combined with bioactive compounds supplementation (*25*). Together with colonization and pharmacological treatment, they can give mechanistic insights on host, microbiome and drug interaction (*490*).

6.1.3. **Fecal microbiota transplantation**

This procedure consists in transplanting a microbiota from a mouse or human donor, to another recipient mouse, and study the effects that produces in the recipient mouse. Germ-free mice are often used as a recipient mouse (*491*). Sometimes, one wants to investigate what happens if a FMT is performed from a healthy mouse to a mouse suffering a disease such as obesity, or vice versa (*492*). Human fecal material transplantation is also used as a proxy to induce a human disease in the recipient mouse, for example, the mouse receiving the microbiota of an obese person (*493*). They are less elegant than small consortia colonizations, but offer the possibility to make strong associations between microbiota and disease. Additionally, FMT has a direct applicability to treat patients with specific diseases, such as recurrent *Clostridium difficile*, with higher rates of complete remission (*494*). Several clinical trials and at least one FDA approved protocol, uses this approach to treat diseases such as obesity, depression or autism(ClinicalTrials.gov Identifier: NCT03281044) (*495*), *(222)*.

6.1.4. **Knock out or in (KO or KI)**

Knock mice allow more mechanistic insights similar to microbial colonization, but aim to study the microbiome response to a mouse directed variability compared to the wild type. For example, to study how the gut microbiota composition changes when a specific host gene is knocked out (*496–498*). Knocks-in experiments can provide valid mechanistic insights when a gene (not previously encoded in the host gene), is introduced to the host, and the gut microbial composition response is measured.

6.1.5. **Antibiotic models**

Analogous to germ-free or colonization experiments, antibiotic treated microbiome models aim to underpin what happens in the host, if the gut microbiota is depleted with antibiotics (*499*). It can be seen as a means to "mimik" and validate results from germ-free experiments. It has also been shown that recovery after an antibiotic course depends on microbial community context, host diet and environmental reservoirs (*500*).

Aims and Objectives

In this PhD thesis I sought to better understand and characterize the communication between the host epigenome, the host metabolism and its symbiont gut microbiota. To study this crosstalk between the host and its microbiome I have found a suitable biological/ecological framework, created a targeted metabolomics method to measure epigenetically relevant metabolites, and measured several omics layers in a germ-free mouse model. The three main objectives of this PhD thesis are:

- To find a suitable biological framework to integrate the crosstalk between the host and its microbiota. The holobiont, the single biological and evolutionary unit formed by the host and its microbiota, and the poorly explored "nutrient-microbe metabolism-host epigenetic axis" of communication, provides me with an optimal biological framework to integrate all host and microbiota molecular layers involved in the crosstalk microbiota-host.

- To create a targeted metabolomics method to measure epigenetically relevant metabolites, including short chain fatty acids (SCFAs). To understand the epigenetic effect of the chemical conversation between the host and its microbiota, I optimized a targeted metabolomics method capable of measuring epigenetically relevant metabolites produced by the host and microbiota, and SCFAs mainly produced by the gut microbiota.

- To study the communication between the host metabolism, host epigenetics marks and gut microbiota from an holobiont perspective. Using a germ-free mice model, with biological sex and having or not having microbiota as experimental factors, we sought to understand how sex and the presence or absence of microbiota interacted and display different omics phenotypes.

In Chapter 1 (section 7), I introduce the holobiont theory which mainly states that the host and its microbiota form an indivisible and single biological and evolutionary. There are many flavors in which the microbiota can influence the host and vice versa, the host influencing the microbiota, and the holobiont concept procures us with an ecological framework where the features of each biont interact and regulate homeostasis. Host and its microbiota communicate at different molecular levels, for example, through metabolites and their metabolisms. In chapter 2 (section 8), I optimized a targeted metabolomics method that permitted us to measure epigenetically relevant metabolites, to understand better how the host microbiota metabolism can affect epigenetic marks in the host and regulate gene expression. Some of these metabolites are exclusively produced by the host microbiota, such as SCFAs, that are potent histone transferase substrates and histone deacetylases inhibitors, making the link between microbiota metabolites and host epigenetic marks. And finally in chapter 3 (section 9), we used the targeted metabolomics method together with the measurement of other molecular layers such as DNA methylation, gene expression, histone post translational modifications or metal abundances in the liver of the mice, to have a better understanding of the mouse holobiont. Analyzing the model including sex and microbiota status has allowed us to discover a strong relationship between the gut microbiota composition and epigenetic signatures in males.

Chapter 1

Chapter 1 has been peer reviewed and published in Frontiers in Genetics as a Perspective article, in a collection of articles at the interface between chromatin and metabolism. This work has been supported by the European Union's Horizon 2020 under the grant agreement No 675610.

7. Epigenetic Regulation at the Interplay Between Gut Microbiota and Host Metabolism

Gut microbiota communities have coevolved for millions of years in a symbiotic relationship with their mammalian hosts. Elucidating and understanding the molecular mechanisms by which microbiota interacts with its host and how this contributes to the homeostasis of the host is crucial. One of these molecular relationships is the so-called chemical crosstalk between microbiota and host metabolisms, including the poorly explored epigenetic regulation of host tissues by the metabolic activity of gut microbiota in response to changes in diet. DNA methylation and histone modifications are epigenetic marks partly regulated by enzymes such as methylases and acetylases, whose activity depend on host and microbiota metabolites that act as substrates and cofactors for these reactions. However, providing a complete mechanistic description of the regulatory interactions between both metabolisms and the impact on the expression of host genes through an epigenetic modulation, remains elusive. This article presents our perspective on how metabolomic, metagenomic, transcriptomic, and epigenomic data can be used to investigate the "microbiota–nutrient metabolism–epigenetics axis." We also discuss the implications and opportunities this knowledge may have for basic and applied science, such as the impact on the way we structure future research, understand, and prevent diseases like type 2 diabetes or obesity.

7.1. Crosstalk Between Host and Gut Microbiota Metabolisms

The human body cohabit with a diverse community of symbiotic microorganisms and their set of genes, collectively known as the microbiome (*501*). The acquisition of the initial microbiome is a dynamic rather than a static process during early life (*109*). A recent estimation of the number of bacterial cells over human cells in our body has reduced the ratio from 10:1 (*502*) to a 1.3:1 (*1*, *503*) . This implies a similar number of bacterial and human cells in and on the human body. However, the human microbiome encodes for at least 100 times more genes than our genome (*2*). Therefore, the corresponding higher functionality of bacterial genes is a key aspect to understand existing metabolic interactions between the host and its microbiota.

The microbiota helps their hosts to digest dietary fiber; produces some important neurotransmitters (*504–506*), hormones, and vitamins (*507*, *508*); helps in training the host immune system (*507*, *509*); and protects against pathogens (*510*), among many other functions. However, unbalanced microbiota can also cause disease. Some common diseases in western societies such as obesity and type 2 diabetes are associated with shifts in the relative abundance of gut bacteria composition and functionality, compared to the ones observed in healthy individuals. The cause of the microbiota imbalance

(dysbiosis) of unhealthy individuals across age and geography has been mainly correlated with dietary habits (*130*, *511–515*). Therefore, the so-called chemical crosstalk between the microbiota and its host has tangible consequences for the physiological state of the host (*516–518*). However, the molecular mechanisms by which microbiota chemically interact with host cells and regulate gene expression remain largely unknown. In this regard, the role that certain host-microbiota derived metabolites may exert on epigenetic events at the DNA, RNA, and histone level needs to be further investigated.

7.2. The "Microbiota–Nutrient Metabolism–Host Epigenetic" Axis

Differences in the microbiota or epigenome in two genetically identical organisms, such as same-sex inbred mice or monozygous twins, can create differences in susceptibilities to diseases including obesity and type 2 diabetes (*519*, *520*). As with gut microbiota, new studies have demonstrated that epigenetic events are highly dynamic, changing in response to nutrient availability (*521*, *522*) or physical exercise (*523*). DNA methyltransferases, DNA hydroxylases, histone acetyltransferases, histone deacetylases, histone methyltransferases, and histone demethylases are enzymes responsible for adding or removing these dynamic epigenetic modifications. In this regard, endogenous metabolites can regulate gene expression through epigenetic events in host cells (*23*). For instance, histone deacetylation regulated by sirtuin family deacetylases is regulated by the NAD^+/NADH ratio, acetyl-CoA, O-acetyl-ADP-ribose, and nicotinamide (*524*, *525*). Whether gut microbiota metabolism is regulating the concentration and/or activity of endogenously produced metabolites by the host remains largely unexplored, and it is only recently that an increasing number of researchers have started to investigate this possibility (*25*, *526*, *527*). Krautkramer et al have demonstrated that microbial colonization regulates global histone acetylation and methylation in multiple host tissues in a diet-dependent manner.

Short-chain fatty acids (SCFAs) are exclusively produced by the microbial fermentation of dietary carbohydrates, and their abundances are regulated by the composition of the microbiota (*94*, *528*). Importantly, SCFAs, particularly butyrate and acetate, produced by the microbiota, inhibit histone deacetylases (Figure 1) (*93*). Increased levels of histone acetylation promote decondensation and relaxation of chromatin, supporting a more transcriptionally active state of chromatin (*529*).

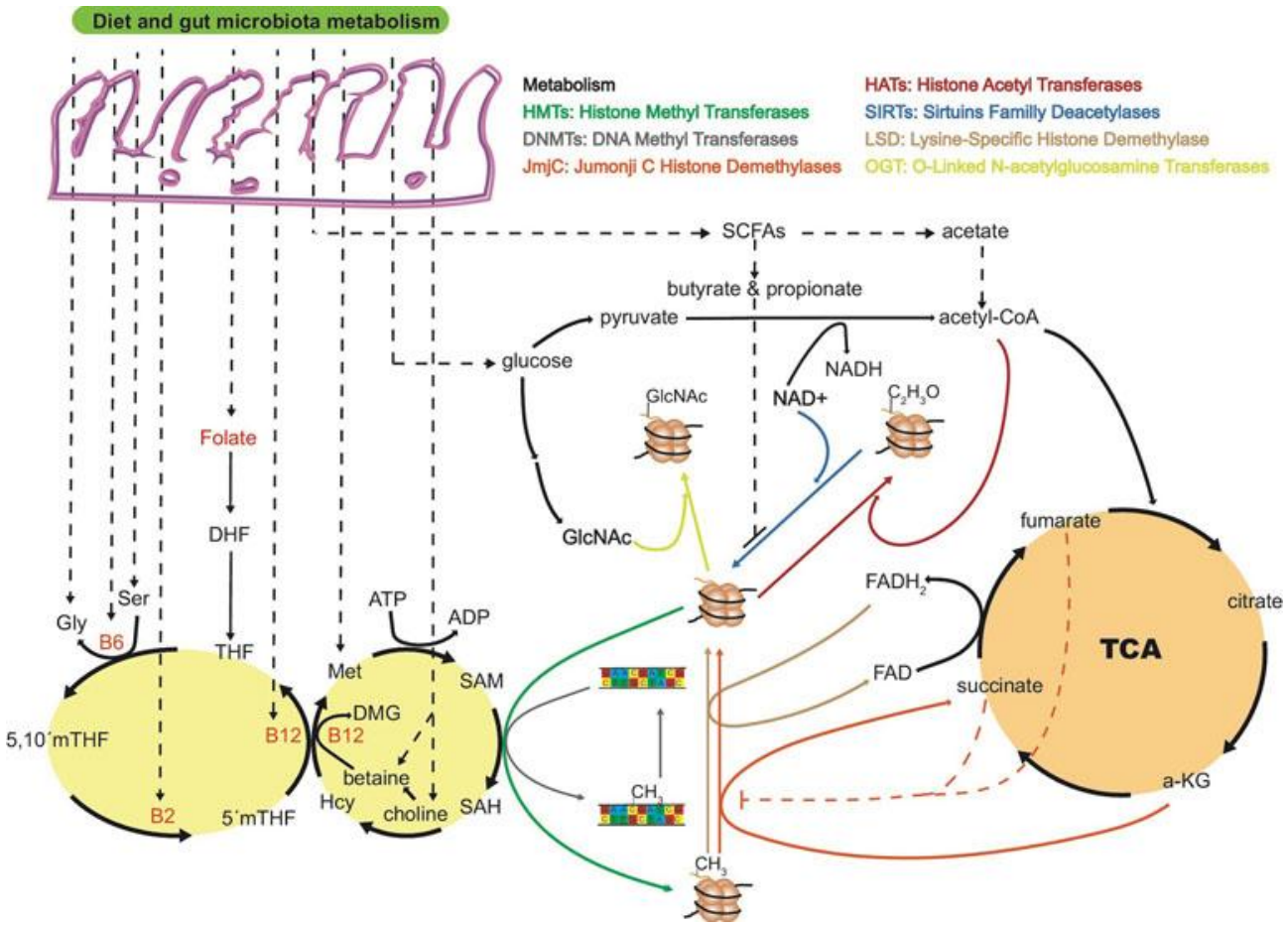


Figure 1. The "microbiota–nutrient metabolism–epigenetics" axis. Most of the key molecules involved in one-carbon metabolism are dietary- and microbiota-dependent, being susceptible to gut dysbiosis or diet intervention. Folate is the precursor of dihydrofolate (DHF) and tetrahydrofolate (THF), and dietary intake is the only source for humans. Together with vitamin B12, 5'methy-THF is in charge of re-methylating homocysteine (Hcy) to methionine (Met), a crucial step in the process of transferring a methyl group to DNA or histones through SAM. The ratio of S-adenosyl homocysteine (SAH) to SAM regulates the overall methylation status of the genome at the DNA or histone level. Vitamins B12, B2, and B6 are key cofactors in the folate cycle that are produced by the microbiota or ingested through diet. Intermediates of the tricarboxylic acid cycle (TCA) are known to positively or negatively regulate histone methylation. For example, alpha-ketoglutarate (α-KG) is known to be an essential substrate for jumonji C histone demethylases (jmjC), and levels of succinate and fumarate can inhibit jmjC demethylases. α-Ketoglutarate is a co-substrate of TET dioxygenases in charge of demethylation processes of histones and DNA. As for jmjC demethylases, increased levels of fumarate and succinate can inhibit TET enzymes with the consequent increased levels of histone and DNA methylation. Short-chain fatty acids (SCFAs) produced by the gut microbiota are also known to inhibit or promote histone PTMs. Butyrate and propionate are inhibitors of sirtuins deacetylases enzymes. Acetate from gut fermentation contributes to the pool of intermediate molecules known to form acetyl-coenzyme A, the major acetyl group donor for histone acetyl transferases (HATs). Acetate is also known to be an inhibitor of histone deacetylases (HDAC), increasing histone acetylation levels and regulating chromatin accessibility. Whether levels of FAD/$FADH_2$, NAD/NADH, TCA intermediates, and other host endogenous epigenetically relevant metabolites are modulated by gut microbiota metabolism needs to be further investigated. DMG, dimethylglycine; ATP, adenosine triphosphate; ADP, adenosine diphosphate; FAD/$FADH_2$, Flavin adenine dinucleotide; NAD^+/NADH, nicotinamide adenine dinucleotide.

Alteration of chromatin state is a possible mechanism by which gut microbiota induces host immune maturation (*530*, *531*). Recognition of a "self" antigen should not only be limited to mammalian host antigens, but also symbiotic microbiota antigens forming part of the whole human ecosystem in a healthy state. The human leukocyte antigen (HLA) or major histocompatibility complex (MHC) gene system encodes many antigen-presenting proteins, which are essential to recognize and distinguish "self" from "non-self" antigens. Colonization of germ-free mice has demonstrated the capacity of microbiota-specific species to activate MHC class II genes (*532*). However, little is known about the immunomodulatory effect of microbial metabolites. A poorly explored possibility is that epigenetically relevant metabolites such as SCFA, highly influenced by microbiota composition, would regulate MHC gene expression by coordinating activity of enzymes that acetylate and methylate histones and DNA allowing chromatin accessibility (*533*).

Regulation of DNA and histone methylation may be driven by complex microbiota–host metabolism interactions involving S-adenosyl methionine (SAM), derived from the essential amino acid methionine through diet (*534*). Folate plays an essential role by re-methylating homocysteine to methionine (Figure 1), thereby ensuring the provision of SAM (*535*, *536*). In this regard, enzymes that are depleted in obese microbiomes are frequently involved in cofactor and vitamin metabolism (*537*), including the production of cobalamin (vitamin B12), pyridoxal phosphate (the active form of vitamin B6), tetrahydrofolate, and folate (*130*, *157*, *507*). Taken together, dysbiosis of microbiota can influence SAM levels and, as a result, alter the methylation status of DNA and histones. Whether dysbiosis of microbiota can alter α-ketoglutarate and succinate levels in specific peripheral host tissues, and regulate the rate of DNA demethylation, is a plausible but little explored possibility. Ten-eleven translocation (TET) enzymes are a key family of DNA and histone demethylases that use α-ketoglutarate as co-substrate. However, due to the structural similarity with α-ketoglutarate, TETs are susceptible to competitive inhibition by fumarate and succinate, causing an increase in histone and DNA methylation levels (*538*).

Another modification that could be regulated by microbial metabolism is histone phosphorylation. In response to a low ATP/AMP ratio indicative of energy status, the AMP-activated protein kinase (AMPK) can translocate to chromatin and phosphorylate histone H2B (*539*). Changes in AMPK activity have been reported in obesity, type 2 diabetes, metabolic syndrome, and cardiovascular disease (*540*). Interestingly, germ-free mice were resistant to obesity and insulin resistance that develop after consuming a Western-style, high-fat, and sugar-rich diet (*541*). The persistently lean phenotype of germ-free animals was associated with increased skeletal muscle and liver levels of phosphorylated AMPK. It is also tempting to speculate that phosphotransferase systems (PTS) overrepresented in the Western diet microbiomes (*511*, *542*) could have an impact on this histone modification.

In short, dysbiosis and reduction of the microbiota diversity can potentially alter the levels of nutrients and metabolites that can potentially act as regulators of DNA methylation and histone modifications either by directly inhibiting enzymes that catalyze the processes, or by altering the availability of substrates necessary for the enzymatic reactions.

7.3. Holobionts, Multifactorial Diseases, and Omic Technologies

Hosts and their microbiota have a very intimate relationship and should be considered as a single biological and evolutionary unit, termed holobiont (*543–546*). In this regard, we could arguably talk about holo–genome, –transcriptome, –proteome, or –metabolome, referring to the combination of both, host and host microbiota molecular layers or modules of information at the DNA, RNA, protein, or metabolite level, respectively (Figure 2). To investigate the dynamics of the holobiont ecosystem network, multi-omic approaches bring unprecedented advantages. Diseases such as obesity and diabetes are known to be multifactorial, and the collection of several -*omic* data (Table 1) from the same holobiont specimen, may provide a detailed molecular description and new mechanistic insights of how dietary nutrients and gut microbiota metabolism can regulate the host phenotype through gene expression and epigenetic and metabolic regulation.

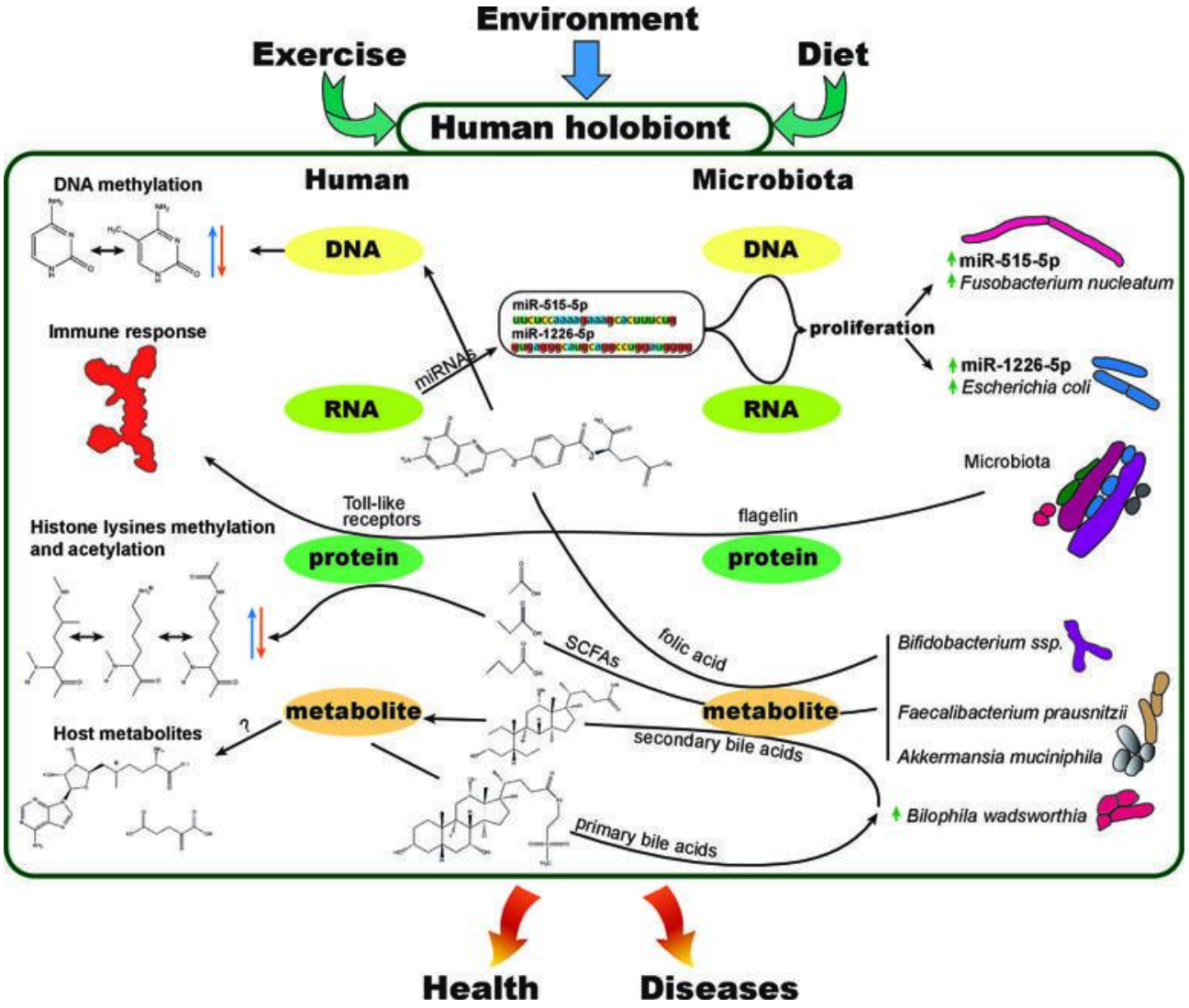


Figure 2 The human holobiont. Representation of few examples of known interactions between different molecular levels within a holobiont. Exercise, environment, and diet can affect the physiology and molecular interactions between human (host) and its microbiota at the DNA, RNA, protein, or metabolite level. As an example, fecal host micro RNAs (miRNAs) are used by the host to modulate the composition of its own gut microbiota, interacting at the microbiota RNA and DNA levels to control microbial growth (*346*). Short-chain

fatty acids (SCFAs), products of gut bacterial anaerobic fermentation of dietary fiber, have been proved to cause changes in histone PTMs in multiple host tissues (*25*). Butyrate is a potent histone deacetylase inhibitor (HDACi), regulating the transcription levels of genes involved in colorectal tumorigenesis (*547*). The direct transformation of dietary nutrients (*516*) and secondary products of host metabolites such as primary bile acids (*548*) evidences the strong interdependence between host and microbiota. Folate production by *Bifidobacterium* spp. is another example of how gut microbiota products can affect epigenetics such as DNA or histone methylation (*549*). Microbiota diversity shifts, products, or bacterial structural components such as flagellin can cause the activation of the immune system as well as impact the immune reconstitution after certain diseases or immunotherapy (*550*). Immune system maturation and allergic disease development are other examples of how the host and its microbiome interact (*551*).

TABLE 1

Omics	Focus	Trait studied	Techniques used	Reference
Epigenomics	DNA modifications	5-methylcytosine	WGBS (whole genome bisulfite Seq)	(Lister et al., 2009)
			RRBS (reduced represented bisulfite Seq)	(Xi et al., 2012)
			MeDIP-Seq (methylated DNA IP Seq)	(Down et al., 2008)
		5-hydroxymethylation	oxBS-Seq (oxidative bisulfite Seq)	(Booth et al., 2012)
		5-formylcytosine	RedBS-Seq (reduced bisulfite Seq)	(Booth et al., 2014)
	RNA modifications	6-methyladenosine	m6A-Seq (m6A specific methylated IP Seq)	(Meyer et al., 2012)
	DNA 3D structure	DNA structure and protein interaction	ChIP-Seq (chromatin IP Seq)	(Barski et al., 2007)
			ATAC-Seq (assay transposase accessible chromatin Seq)	(Buenrostro et al., 2013)
			Hi-C (chromatin conformation capture)	(Lieberman-Aiden et al., 2009)
			DNase-Seq (DNase I hypersensitive sites Seq)	(Boyle et al., 2008)
	RNA transcripts	Transcribed DNA	RNA-Seq (mRNA/size/strand Seq)	(Mortazavi et al., 2008)
			GRO-Seq (global run-on-sequencing Seq)	(Core et al., 2008)
			NET-Seq (native elongating transcript Seq)	(Churchman and Weissman, 2011)
			UMI method (unique molecular identifiers)	(Kivioja et al., 2012)
Metagenomics	Marker gene	Hypervariable region	16S gene (16S amplicon PCR/sequencing)	(Woese and Fox, 1977)
	Whole metagenome	Whole genome	DNA-Seq (regular DNA Seq)	(Tyson et al., 2004)
	Metatranscriptome	RNA	RNA-Seq (regular RNA Seq)	(Gilbert et al., 2008)
Metabolomics	Targeted	Known metabolites	QqQ (triple quadrupole)	(Lu et al., 2008)
	Untargeted profiling	Unknown metabolites	qTOF-MS (quadrupole time of flight)	(Patti et al., 2012)
			Orbitrap-MS	
			NMR (nuclear magnetic resonance)	(Brindle et al., 2002)
Proteomics	Histones PTMs	H2A, H2B, H3, and H4 modifications	Bottom-up	(Sidoli et al., 2012)
			Middle-down	
			Top-down	
			MALDI-imaging mass spectrometry	(Lahiri et al., 2016)

The integration and combination of these techniques have the potential to reveal more mechanistic insights on how gut microbiota influence epigenetics and gene expression, ultimately affecting host health. IP, immune precipitation; Seq, sequencing; m6A, 6'methyl adenosine; MS, mass spectrometry; LC, liquid chromatography; GC, gas chromatography; PTMs, posttranslational modifications; MALDI, Matrix-assisted laser desorption ionization.

Table 1. Popular omic techniques. Popular omics techniques in the fields of epigenomics, metagenomics, metabolomics, and proteomics.

To study a complex metabolic disorder such as familial type 1 diabetes mellitus (T1D), (*552*) Heintz-Buschart et al. used a combination of host genomics and proteomics together with metagenomics, metatranscriptomics, and metaproteomics to demonstrate a pronounced family membership effect in the structuration and functionality of the microbiomes. They observed a correlation between certain human pancreatic enzymes and the expression of specific microbial genes involved in key T1D metabolic transformations.(*25*) Krautkramer et al. used a combination of metabolomic, proteomics of histones, transcriptomic, and metagenomic techniques in conventional and germ-free mice, to demonstrate how SCFAs produced by the microbiota, or supplemented exogenously to germ-free mice, regulate histone post-translational modifications (PTMs). Comparing histone PTMs and transcriptional profiles between conventional, germ-free mice and germ-free mice supplemented with SCFAs, Krautkramer et al. concluded that SCFAs alone are partially causative for histone PTMs. In the same direction, (*40*) Thaiss et al. used a combination of metagenomics, transcriptomics,

epigenomics, and metabolomics together with imaging electron microscopy, to identify a diurnal rhythmicity in the microbial biogeography, metabolic profile, and metagenomic functionality as critical orchestrators of host epigenetic marks and gene expression. Thaiss et al. have demonstrated that host epigenetic and transcriptional circadian oscillations are partially dependent on environmental signals such as microbiome metabolite dynamics in the intestines and that peripheral organs "sense and adapt" to this circadian metabolite rhythms in a similar manner.

Overall, multi-omic approaches will facilitate the structuration of future research, improve patient stratification toward a more personalized care, and open new avenues to evaluate the effectiveness of functional probiotics, functional foods, or nutritional interventions aimed at regulating host gene expression in health and disease.

7.4. Opportunities for Biomedical and Clinical Research

The development and improvement of new technologies and bioinformatic tools are advancing biomedical research at a fast pace. Multi-*omic* experiments are allowing researchers to obtain mechanistic insights on the human holobiont homeostasis, improving decision-making for next experiments to be performed. In a research context, extensive collection of -*omic* data will allow integration of information into modulable and controlled models of microbial communities (*553*). However, it is important to identify confounding factors in longitudinal -*omic* studies. Standardization of methods and techniques to reduce noise and bias in microbiome research will improve how we translate lab findings into the clinic (*449*).

Using *germ-free* or gnotobiotic mouse models provide a framework to manipulate the gut microbial composition in a controlled manner. These models can be used to study the chemical crosstalk between host and microbiota by uncovering specific epigenetic changes in host cells induced by colonization of specific bacterial strains. Colonizing gnotobiotic mice with single strains or small bacterial consortia, instead of whole fecal transplants, might bring more accurate information to understand specific molecular mechanisms and molecular pathways involved in the host epigenetic regulations. Microbiota-induced dysbiosis with antibiotics might provide a useful approach to validate findings in gnotobiotic models by partially mimicking the effects of absence of microbiota or a reduction in the microbial diversity. Interestingly, organoids might also provide a challenging but very useful *in vitro* culturing system to study host–microbiota interactions (*554*).

In a clinical context, gaining knowledge on the "microbiota–nutrient metabolism–host epigenetics axis" using multiple -*omic* approaches in combination with microbiota modulation therapies, has the potential to prevent and treat more efficiently metabolic diseases:

1. Fecal microbiota transplantation (FMT): The use of microbiota modulation therapies such as FMT to treat recurrent *Clostridium difficile* infections has been proved significantly more efficient than a vancomycin treatment (*555*, *556*). Recently, the potential of FMT as a microbiome modulation technique for treating metabolic (*557*, *558*), neurological (*559*), and immunological disorders (*560*) has been tested, improving these conditions by partially restoring microbiota diversity and functionality. Assessing long-term host epigenetic effects of FMT has to be further investigated.

2. Microbiota as drug target: The unique microbiome composition of each person is probably responsible for different susceptibilities to the same nutrient, pollutant, or drug treatment (*561*). Microbiota-derived metabolites can enter the bloodstream and interact with drug treatments, impacting the efficacy, toxicity, and clearance of the drug. Diagnosing or treating diseases using microbiota-targeted drugs, probiotics, and use of bacteriophages or engineered bacteria has recently re-emerged (*562–565*).

3. Nutritional intervention, probiotics, and prebiotics: The acute consumption of a cocktail of probiotics containing a selection of five strains of *Lactobacillus* and five strains of *Enterococcus* modulates the microbiome and enhances SCFAs production in human and mice (*566*). The consumption of some probiotics has proven to be beneficial and ameliorates stress felt in healthy women (*567*), improves insulin sensitivity (*568*), protects against infections (*569*), and helps to restore microbiota after distortion by antibiotics (*570*), among other benefits. Promoting SCFAs producing bacteria in the host gut by a nutritional intervention that increases fiber consumption (*571*) may have an epigenetic effect in the host (*25*).

7.5. Conclusions and Perspective

The study of the "microbiota–nutrient metabolism–host epigenetic" axis has great potential to reveal the molecular mechanisms by which gut microbiota composition affects the expression of genes in their hosts. In a human holobiont context, this axis is relevant to understand, prevent, diagnose, and treat the existing epidemic of metabolic disorders such as type 2 diabetes and obesity. Microbiota is a key player in health outcomes due to the potential myriad of metabolites that can produce and interact with any cell of our body through systemic circulation. Those metabolites are coming from direct transformations of nutrients available in the gut microbiota or from secondary transformed host products. The link between epigenetic marks and gut microbes appears to be mediated by host-microbial metabolites that act as substrates and cofactors for key epigenetic enzymes in the host. A disruption in the composition of the gut microbiota may lead to unbalanced key metabolites that sequentially may impact epigenetic pathways and alter gene expression. The implementation of multi-omic approaches to study the human holobiont will facilitate the stratification of patients toward a personalized-oriented care, improving disease prevention, diagnostics, drug election, and treatment efficiency.

7.6. Contributions

The idea of the nutrient-microbe metabolism-host epigenetic axis was conceived by my supervisor Oscar Yanes and I contributed to find a suitable biological framework where to contextualize the underlying biology of this new proposed axis of communication. I drew all the figures, created the tables and created the first draft of the manuscript, and together with Oscar we created the final version of the manuscript.

Chapter 2

Chapter 2 has been submitted to the Journal of Biological Chemistry. This work has been supported by the European Union's Horizon 2020 under the grant agreement No 675610.

8. A targeted metabolomics method to detect epigenetically relevant metabolites

Metabolites play a central role in the chemical crosstalk between metabolism and epigenetics. Epigenetically relevant metabolites are substrates, products and cofactors that can act as activators or inhibitors of epigenetic enzymes, which control gene expression by adding or removing chemical marks in the DNA, RNA and histones. Diet composition, and biosynthetic pathways encoded in the gut microbiome and the host genome are the main sources of these metabolites for mammals. Despite the increasing interest in the study of the 'microbiota-nutrient metabolism-host epigenetic axis' to understand health and disease, there is a lack of a sensitive and easy analytical method to detect epigenetically relevant metabolites simultaneously. Here, we show an straightforward biphasic extraction where the organic phase is directly analyzed by GC-EI MS to detect short-chain fatty acids and formate without chemical derivatization, and the aqueous phase is analyzed by HILIC coupled to ESI-MS/MS, which together can cover >30 epigenetically relevant metabolites in biological samples such as liver, plasma or feces. In addition, we propose a stable isotope tracing method based on multiple-reaction monitoring (MRM) transitions by LC-QqQ MS to understand how 13C-labeled glucose or glutamine are used to build SAM, acetyl-CoA and UDP-GlcNAc, the main methyl, acetyl and glucosyl group donors in epigenetic modifications, respectively. We anticipate that our method will complement epigenomic and proteomic analyses adding another layer of molecular information towards mechanistic insights.

8.1. Introduction

Eukaryotes, and especially mammals, have developed a chemical crosstalk between metabolism and the reversible chemical signatures in their DNA, RNA and histone proteins, responding to environmental changes (*572*) Known as epigenetic marks, such signatures can regulate the function, accessibility or 3D architecture of DNA, RNA and histones (*317*, *353*, *573–576*). The crosstalk between metabolism and epigenetics occurs through specific metabolites that are responsible for regulating the enzymes that add or remove chemical moieties such as methylations, acetylations or glycosylations in these biopolymers (*572*).

The host genome encodes enzymes that produce the tricarboxylic acid (TCA) cycle intermediates succinate and alpha-ketoglutaric acid (αKG), that can act as inhibitor and co-substrate of Jumonji demethylases, respectively (*23*, *577*), or fumarate, capable of modulating histone and DNA methyltransferase activity (*578*). Other examples of host synthesized metabolites are S-adenosylmethionine (SAM), a key metabolite in one carbon metabolism and the main methyl group donor in methylation reactions (*579*); nicotinamide adenine dinucleotide (NAD^+), which mediates many redox reactions and controls deacetylase activity of sirtuins (*580*); or acetyl coenzyme A (acetyl-CoA),

the principal acetyl group donor and the preferred substrate for histone acetyltransferases (HATs) (*581*). However, other epigenetically relevant metabolites can only be ingested through the diet or produced by the gut microbiota (*25*, *56*). This is the case of methionine, an intermediate of the one carbon metabolism involved in the production of SAM, and as an essential amino acid, it needs to be ingested (*56*). In addition to dietary nutrients, the gut microbiota has been recognized as a source of folate, whose levels are associated with global DNA methylation status (*582*). Similarly in its origin, butyric acid is a short-chain fatty acid (SCFA) mainly produced by the gut microbiota (*25*) that acts as an acetyl group donor and a potent histone deacetylase (HDAC) inhibitor.

Despite the development of specific methods based on LC-MS/MS and GC-MS to detect and quantify epigenetically relevant metabolites individually (e.g, SAM, folate, acetyl-CoA, vitamin B12, butyric acid) (*583–587*) or as small sets of related compounds such as folic acid derivatives (*588*) or SCFAs (*589*), the large structural heterogeneity and physico-chemical diversity of these metabolites has hampered the development of a single and comprehensive method that covers the most important metabolites involved in epigenetic reactions.

Here, we show the optimization of a metabolite extraction protocol with no chemical derivatization combining complementary GC-MS and LC-MS/MS technologies targeting >30 epigenetically relevant metabolites, including intermediates of the TCA cycle (*590*), purine and pyrimidine synthesis (*591*), vitamins and coenzymes (*581*, *592*), one carbon metabolism (*79*), essential and non-essential amino acids (*593*) and SCFAs (*92*). In addition, we propose a targeted metabolomic approach based on multiple reaction monitoring (MRM) transitions to track the fate of the stable isotope ^{13}C from uniformly labeled glucose or glutamine in the synthesis of SAM, acetyl-CoA and UDP-GlcNAc, the main methyl, acetyl and glucosyl group donors in epigenetic modifications, respectively.

8.2. Result

8.2.1. Optimizing extraction, chromatographic and mass spectrometry conditions.

Due to the chemical heterogeneity of epigenetically relevant metabolites (Supplementary table 1), we explored eight metabolite extraction methods, including biphasic and monophasic extractions (see Methods section for details) coupled to reversed-phase (RP-C18), hydrophilic interaction liquid chromatography (HILIC) or gas chromatography (GC) to improve separation, and MRM transitions for detection by mass spectrometry. Optimization was performed using 42 standard compounds covering the microbiota-nutrient host metabolism-epigenetic axis (Figure 1) (*22*), which were subsequently monitored in liver tissue and cecal content from mice, and human serum.

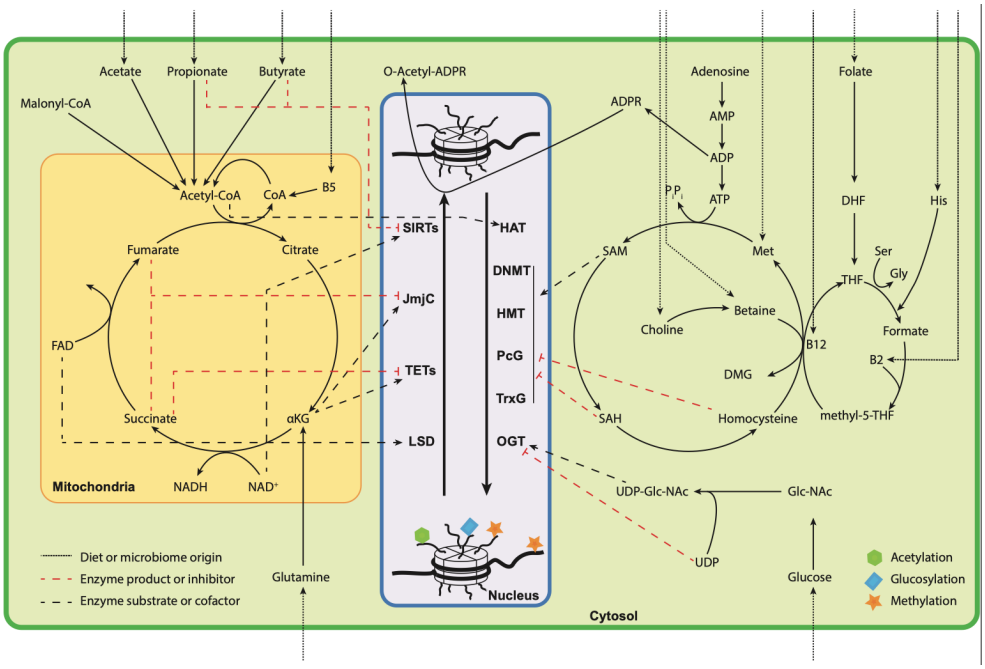


Figure 1. Microbiota-Nutrient host metabolism-epigenetic axis, chemical crosstalk between metabolism and epigenetics. Schematic representation of how epigenetically relevant metabolites regulate epigenetic writers and erasers, connecting microbiota, metabolism and epigenetic marks. CoA: coenzyme A; B5: pantothenic acid; FAD: flavin adenine dinucleotide; αKG: alpha-ketoglutarate; NAD+/NADH: nicotinamide adenine dinucleotide/reduced form; AMP: adenosine-5'-monophosphate; ADP: adenosine-5'-diphosphate; ATP: adenosine-5'-triphosphate; ADPR: ADP-ribose; DHF: dihydrofolate; THF: tetrahydrofolate; His: histidine; Ser: serine; Met: methionine; Gly: glycine; DMG: dimethylglycine; SAM: S'-adenosylmethionine; SAH: S'-adenosylhomocysteine; B12: cyanocobalamin; GlcNAc: N'-acetylglucosamine; UDP-GlcNAc: uridine-diphosphate-N-acetylglucosamine; UDP: uridine-diphosphate; SIRTs: sirtuins (deacetylases); JmjC: Jumonji C (demethylase); TETs: ten-eleven translocation methylcytosine dioxygenases (demethylases); LSD: lysine-specific histone demethylases; HAT: histone acetyl transferases; DNMT: DNA methyltransferases; PcG: polycomb group (methylases); TrxG: Trithorax-group proteins (methylases); OGT: O-GlcNAc transferase.

A biphasic extraction adapted from *Lotti et al.* (*589*) consisting in an acidic aqueous solution (with phosphoric acid at 15%) combined with the organic solvent methyl tert-butyl ether *(MTBE)* resulted in the best extraction method. The *Phospho/MTBE* extraction was able to efficiently extract free SCFAs and formate in the organic phase, while dissolving other relevant epigenetic metabolites in the polar acidic phase. When compared with the best monophasic extraction using *ACN/MeOH/H_2O* we observed a high degree of overlap (Figure 2a), however the *Phospho/MTBE* was the only extraction method capable of recovering SCFAs and formate from every biological matrix tested, that is, liver tissue, caecum and serum (Figure 2b-c,e).

Historically, free SCFAs and formate have been a challenging set of compounds to analyze by liquid and gas chromatography coupled to mass spectrometry, requiring chemical derivatization to stabilize them ((*25*, *594*)). When we compared the *Phospho/MTBE* method with another method that required chemical derivatization (*595*), we observed that the derivatization step introduced high background noise in the GC-EI MS spectra, particularly for matrices where the concentration of formate and the SCFAs was low (e.g., serum) (Figure 2d). In contrast, the organic phase of the *Phospho/MTBE* method analyzed directly by GC-EI MS without chemical derivatization produced cleaner spectra, facilitating

the detection and quantification of free SCFAs and formate in liver tissue and serum (Figure 2b,e), where their concentrations are significantly lower than in the gut content (i.e., caecum, Figure 2c) (*596*).

In parallel, we optimized the chromatographic conditions to analyze the acidic aqueous phase of the *Phospho/MTBE* biphasic extraction. We tested three commonly used stationary phases in metabolomics: (i) reversed phase C18; (ii) hydrophilic interaction liquid chromatographic (HILIC) with ethylene bridged hybrid coated particles (BEH); and (iii) a HILIC zwitterionic stationary phase (HILIC-z). The HILIC-z column performed better than RP-C18 and HILIC, retaining and improving the elution peak shape of most epigenetically relevant metabolites (supplementary figure 2a-d), particularly those phosphorylated, assisted by the additive medronic acid in the mobile phase (supplementary figure 2a-b and d). In addition, we tested another zwitterionic column ZIC-pHILIC (see methods section), however it resulted in broader peaks compared with HILIC-z (supplementary Figure 3a-d).

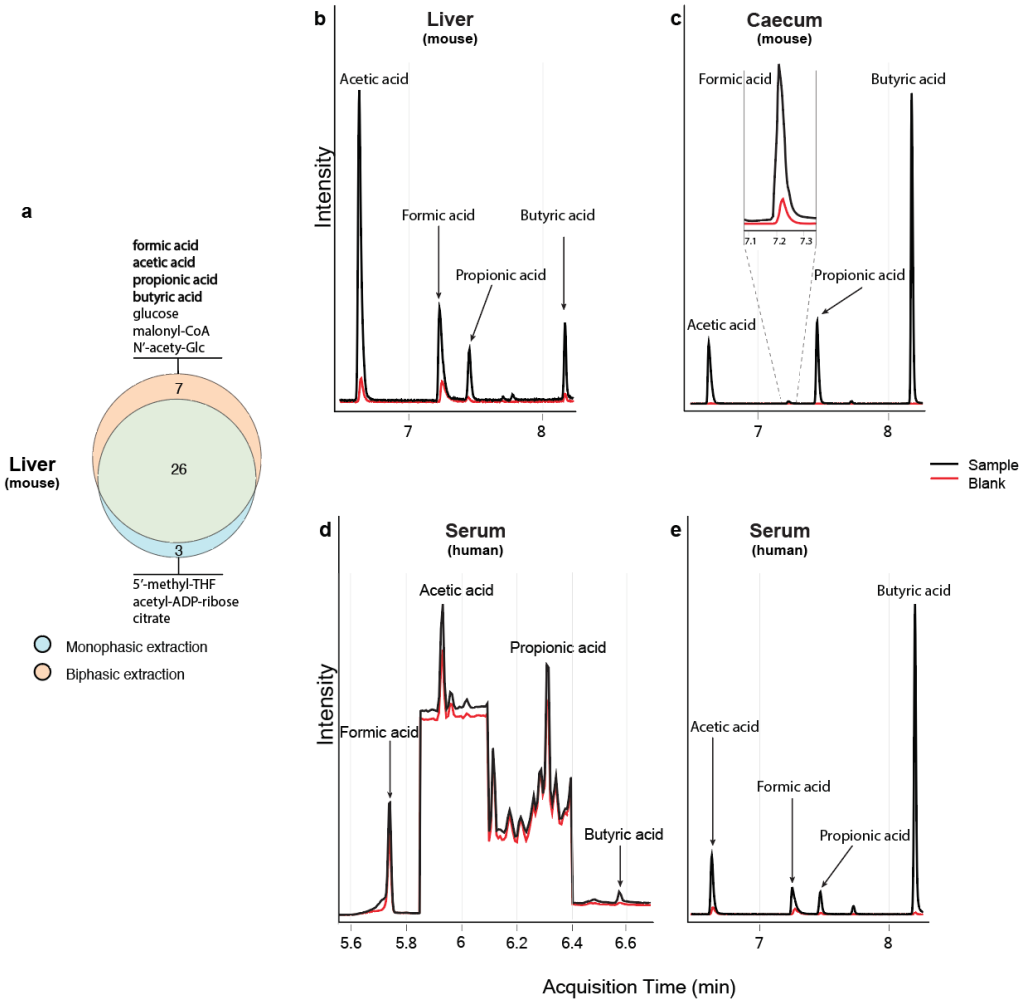


Figure 2. Epigenetically relevant metabolites extracted with a mono- and bi-phasic extraction from liver tissue. (a) Venn diagram showing the overlap of extracted metabolites from liver between the monophasic and biphasic extraction protocols. A complete list of extracted metabolites for each matrix and extraction method can be found in supplementary table 1-2. (b-c) Total ion chromatogram (TIC) of extracted SCFAs and formic acid from mouse liver and caecum content respectively (black line), compared to the corresponding blank sample (red line). (d) TIC of extracted SCFAs and formic acid from human serum using chemical derivatization (black line) and the corresponding blank sample (red line). (e) TIC of extracted SCFAs and formic acid from human serum using the biphasic extraction without chemical derivatization.

8.2.2. **SCFAs and formic acid measured in germ-free (GF) and conventional (CV) mice.**

The main advantage of the method presented herein is the ease of measuring SCFAs (acetic, butyric, propionic and formic acid) as well as other epigenetically relevant metabolites with a single metabolite extraction and without using chemical derivatization. To prove that our method can be used to dissect the chemical crosstalk between metabolism and epigenetics –including the gut microbiota metabolic activity– we tested the method in three different matrices: liver and caecum content from germ-free (GF, n=10) and conventional (CV, n=10) mice, and human serum (n=3). GF mice are specially-raised animals devoid of all microorganisms.

As expected, we detected significantly lower levels of SCFAs in the caecum of GF mice compared to the CV group (Figure 3, left panel). In contrast, the amount of SCFA in the liver of GF and CV mice was similar (Figure 3, right panel), which suggest that GF animals compensate for the absence of microbiota by increasing the activity of pathways that generate SCFAs in the liver, such as beta-oxidation, glycolysis or amino acid degradation. Interestingly, free formate levels in cecum exhibit small but significantly lower levels in GF compared to CV mice, indicating that the cecum microbiota is not the major producer of this important metabolite of the one-carbon metabolism (*597*, *598*).

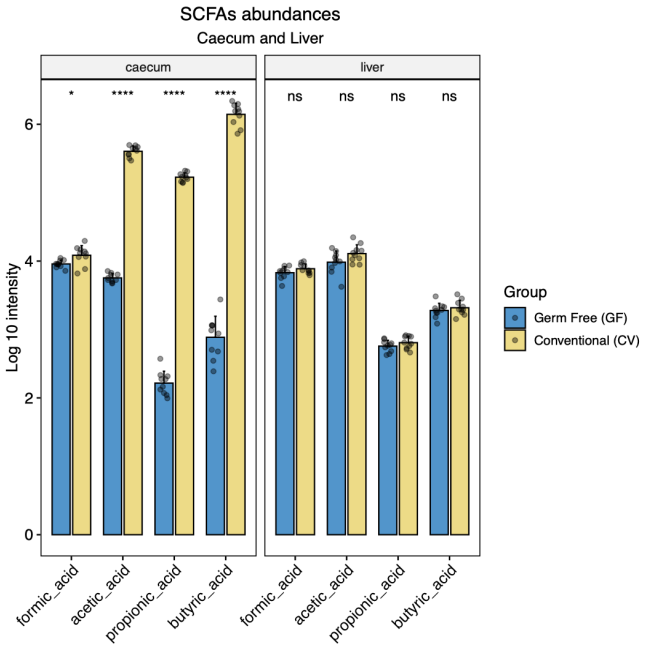


Figure 3. SCFAs in germ-free (GF) and conventional (CV) mice. Abundances in caecum content (left panel) and liver (right panel) of formic acid and short chain fatty acids (SCFAs) in germ-free GF (n=10) and conventional CV (n=10) mice. (Wilcoxon tests, BH adjusted p values: ns: p > 0.05; * p <= 0.05; **: p <= 0.01; ***: p <= 0.001; ****: p <= 0.0001)

8.2.3. **Tracking the fate of [U-^{13}C]glucose in the synthesis of SAM, acetyl-CoA and UDP-GlcNAc**

The deliberate introduction of a stable isotope into a compound perturbs the isotopologue and isotopomer distribution, and therefore, its mass spectrum. This has been extensively exploited using labeled compounds as tracers for the study of physiological and metabolic processes. However, classical stable isotopic labeling studies are mostly based on MS1 isotopologue distributions (*412*), which only provides information about the number of atoms labeled, lacking positional (i.e., structural) labeling information.

Based on our optimized extraction and chromatographic conditions for epigenetically relevant metabolites, we propose a novel strategy for quantification of positional isotopomers using MRM transitions in QqQ MS. Our approach feeds from existing information of biosynthetic pathways and the fragmentation pattern of metabolites of interest to design specific MRM transitions that monitor all possible (labeled and unlabeled) isotopomers, determining the exact position and metabolic origin of the labeled ^{13}C atoms in the *de novo* synthesis of SAM (Figure 4b), acetyl-CoA and UDP-GlcNAc (Supplementary figure 4-6), the main methyl, acetyl and glucosyl group donors in epigenetic modifications, respectively.

SAM was analyzed in mouse embryonic fibroblasts (MEF) at two time points (6h and 72h) as described in Kovatcheva M. et al. (preprint 2023). The proposed MRM transitions monitor all biosynthetic combinations contributed by glucose carbons, which includes the ribose moiety through the pentose phosphate pathway (PPP), the purine nucleotide through *de novo* synthesis of glycine (C4 and C5) and folate cycle (C2 and C8), and the methyl group from methionine. Our data indicates that glucose metabolism contributes to the synthesis of SAM in a time and moiety dependent manner (Figure 4a). After 72h in culture with [U-^{13}C]glucose, the most abundant isotopomer is SAM +5 corresponding to the ^{13}C-enrichment of the ribose moiety. Interestingly, the activity of the folate cycle makes a greater contribution to the *de novo* purine nucleotide synthesis by introducing carbon 8 (C8) and/or carbon 2 (C2) into the purine ring, than the remethylation pathway from homocysteine to methionine and back to SAM. Overall, our approach shows the high number of labeled isotopomers in SAM and their percentages within the total pool of SAM, which would not be distinguished by analyzing isotopologues from MS1 data (Figure 4a).

Acetyl-CoA was analyzed in U-87 MG cells isolated from malignant gliomas. Acetyl-CoA has three chemical moieties originating from glucose metabolism (Supplementary Figure 4b): adenine C4 and C5 from glycine and adenine C2 and C8 from folate cycle, the ribose from the pentose phosphate pathway, and the acetyl group from pyruvate through glycolysis (named A, B and C respectively in supplementary Figure 4 and 5). Our data indicates that glucose metabolism contributes with carbon atoms needed for synthesizing acetyl-CoA and SAM in a time and moiety dependent manner (Figure 4.a, and supplementary 4a). In the ribose moiety B (M+5) of acetyl-CoA, we observe that glucose is the main source of labeled carbons in both time points. At 72h, cells have incorporated nearly twice as much amounts of labeled carbon atoms in comparison to the first time point in both acetyl-CoA and SAM. Meaning that more labeled atoms from the pentose-phosphate pathway (PPP) have been incorporated into the ribose of acetyl-CoA and SAM. PPP is a crucial metabolic route that produces

nucleotides necessary along the cell cycle such as ribose-compounds originated from glucose (*599*). Our measurements indicate a preferential use of glucose over glutamine derived carbons to synthesize ribose in acetyl-CoA (moiety B, Supplementary figure 4a). In the adenine, moiety A of SAM and acetyl-CoA, we observe a mild incorporation of labeled carbon atoms at the first time point with a small increase in only labeled adenine for acetyl-CoA in the second time-point. This slower rate of incorporation of labeled carbon atoms into the adenine moiety (A), compared to a faster incorporation of labeled carbon atoms into the ribose moiety (B) can be explained by two facts: 1) the biosynthesis of adenine (purine ring) is build up one or few atoms at a time and ribose is attached through the process (*599*); 2) the ribose is synthesized in the PPP as result of fewer oxidation reactions of the glucose 6-phosphate (glucose phosphorylated form), being ribose metabolically faster to synthesize than adenine (*599*). In addition, in the acetyl-CoA molecule, carbon atoms are incorporated by the transfer of adenosine from ATP to phosphorylated pantothenate, where adenosine is a combination of the adenine (A) and ribose (B) moieties (M+7). However, for acetyl-CoA, the signal for both moieties being labeled at the same time (M+7) is around a third of the signal of only the ribose moiety (M+5) being labeled, showing more labeled atoms at 72h as for in moiety B. This observation indicates that cells are actively incorporating carbones from glucose in a faster manner to the ribose (M+5), than to the ribose and adenine together (M+7) or adenine (M+2) alone in the acetyl-CoA molecule. In the SAM molecule, the incorporation rate of labeled carbon atoms in the adenine and ribose (M+7) at the same time is very small compared to the acetyl-CoA molecule, and the main labeled atoms in the SAM molecule come from incorporation of labeled atoms from the ribose (M+5).

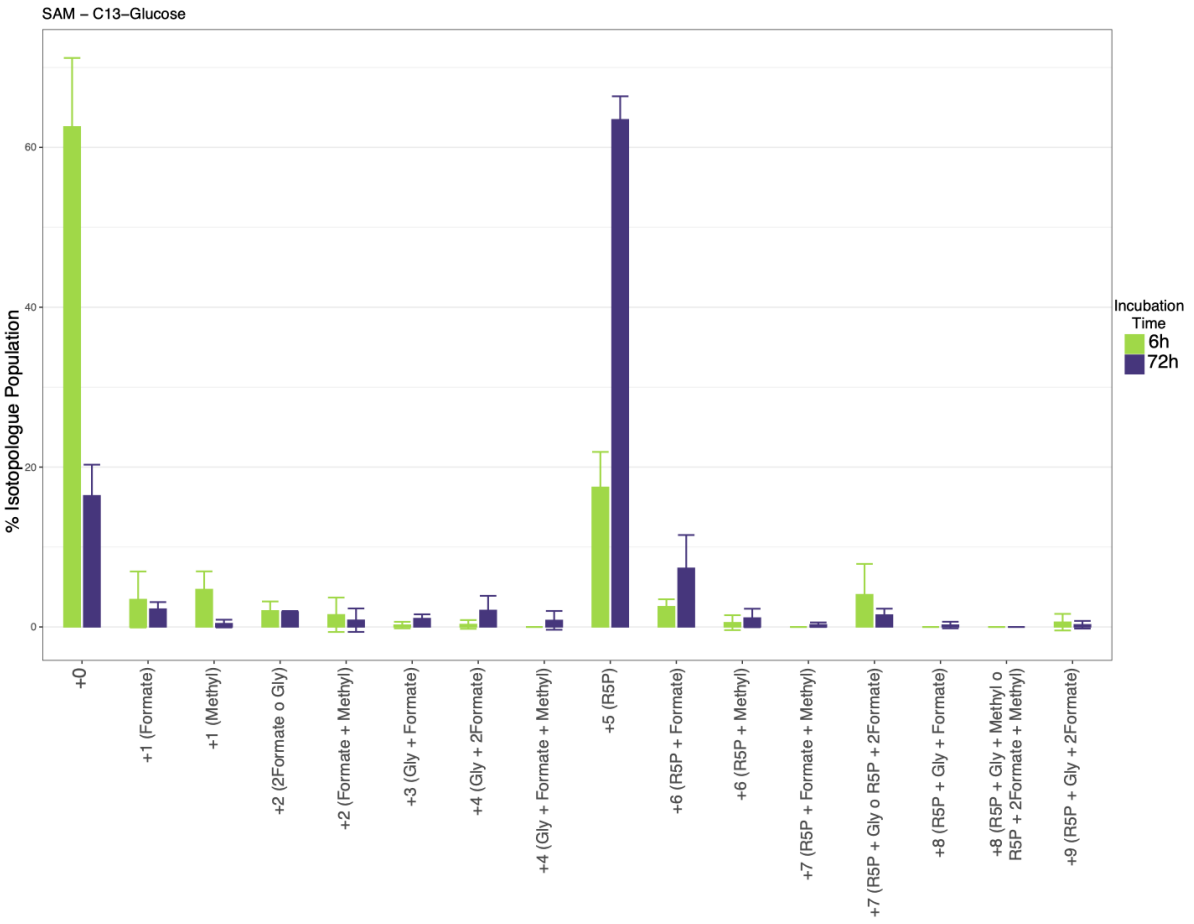
In the acetyl group, moiety C (M+2) of acetyl-CoA, we observe higher labeling rate at 20h compared to 72h when using glucose as source of labeled ^{13}C atoms. We suspect that at the later time point, we are observing lower levels of labeling in the acetyl group (C) as a result of the transference of the acetyl group to histone proteins to promote the activation of growth genes (*528*) or carboxylating the acetyl group to malonyl-CoA which is the first and essential step in de novo fatty acid synthesis occurring during cell proliferation (*600*). In contrast, glutamine contributes very little to the different carbon atoms present in the moieties A and B of the acetyl-CoA molecule. Only around 10% of the labeled atoms were incorporated from labeled glutamine to the acetyl group of the acetyl-CoA molecule, ending up forming part of the moiety C (M+2) in similar proportions in both time points. Biologically, glucose seems to have an initial role to provide acetyl groups to transfer in the acetyl-CoA molecule, and glutamine contribute to a steady flow of carbons to the acetyl group of the acetyl-CoA molecule, ensuring availability of acetyl groups to be transferred.

In the methyl group, moiety C (M+1) of the SAM molecule, we observe a very small incorporation of labeled carbon atoms at 72h after subtracting the natural isotopic distribution. But taking together all the possible combinations of the methyl group moiety (C) with the other two moieties (A and B), we observe a small increase in labeled moieties containing the methyl group. This observation suggests that a small rate of the methyl group is incorporated from labeled glucose through the glycine and serine pathway in contrast to what has been suggested recently (*601*).

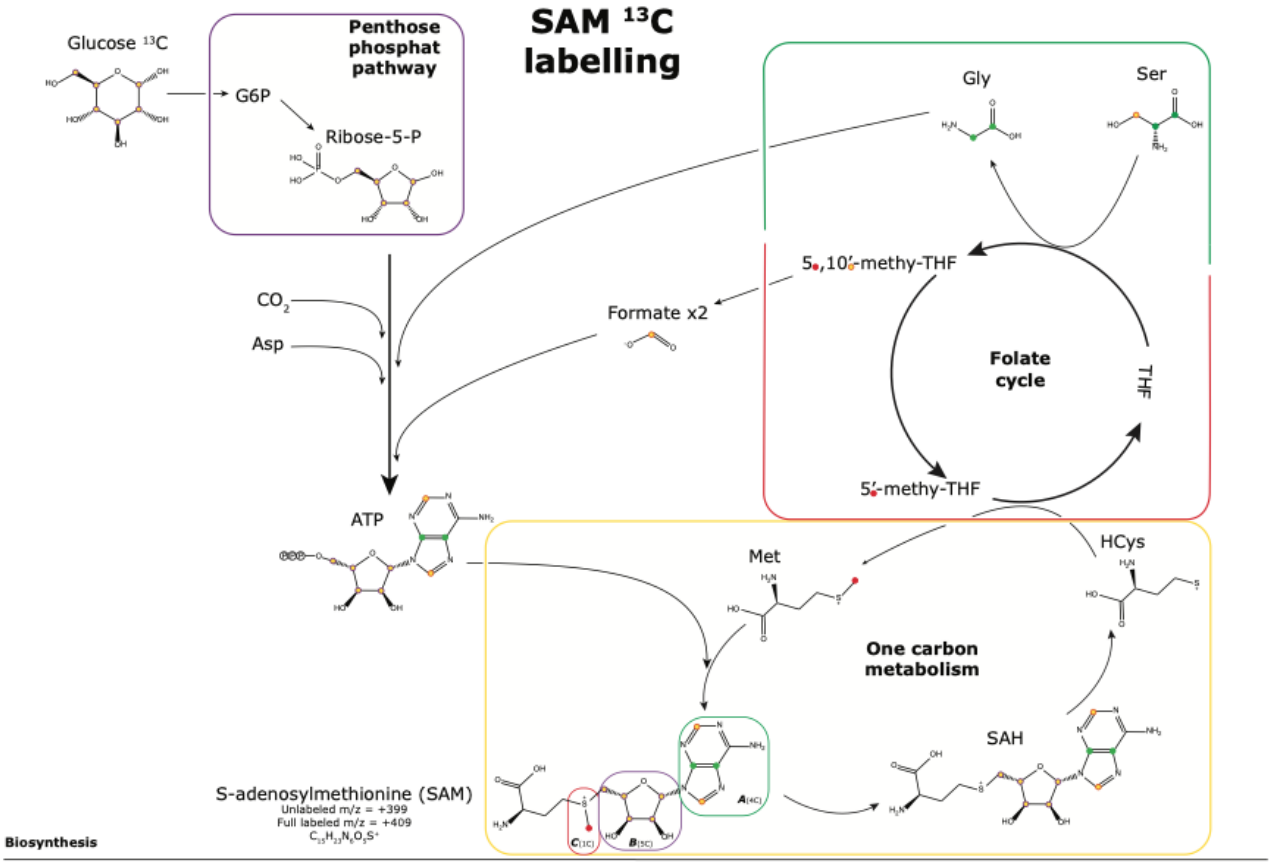
We also performed the same methodological routine described for acetyl-CoA and SAM for UDP-N-acetylglucosamine (UDP-GlcNAc), a metabolite known to be the source of glucose in post translational modification of histones, so-called histone glycosylation. The results point in the same direction as for acetyl-CoA and SAM, the ribose moiety in the UDP-GlcNAc molecule is the moiety

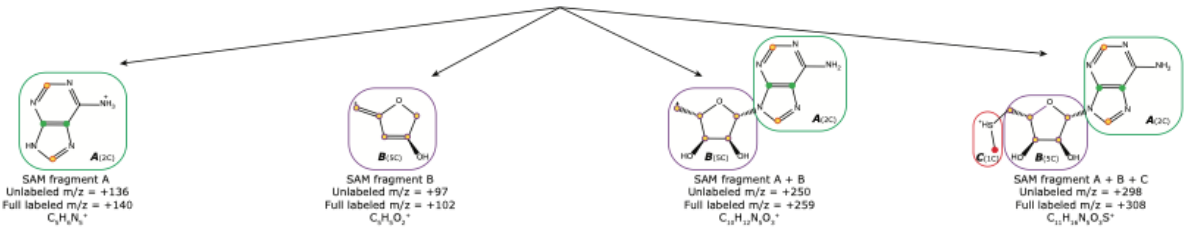
that contributes the most to incorporate labeled carbons (supplementary Figure 5). In this experiment, we had cultured cells only exposed to labeled glucose for 72h hours but not to labeled glutamine, nor 20h of incubation (Supplementary methods).

a



b





Labeled moiety	Parent ion m/z	Labelling origin	Transition m/z	Monitored ^{13}C trans
Unlabeled	399	--	136	--
A	400	+1 (For)	137	A
C	400	+1 (Me)	136	--
A	401	+2 (2For or Gly)	138	A
A + C	401	+2 (For+Me)	137	A
A	402	+3 (Gly+For)	139	A
A + C	402	+3 (2For or Gly+Me)	138	--
A	403	+4 (Gly+2For)	140	A
A + C	403	+4 (Gly+For +Me)	139	A
B	404	+5 (R5P)	136	--
A + C	404	+5 (Gly+2For +Me)	140	A
A + B	405	+6 (R5P+For)	137	A
B + C	405	+6 (R5P+Me)	136	--
A + B	406	+7 (R5P+Gly or 2For)	138	A
A + B + C	406	+7 (R5P+For +Me)	137	A
A + B + C	407	+8 (R5P+Gly or 2For+Me)	138	A
A + B	407	+8 (R5P+Gly +For)	139	A
A + B	408	+9 (R5P+Gly +2For)	140	A
A + B + C	408	+9 (R5P+Gly +For+Me)	139	A
A + B + C	409	+10 (R5P+Gly +2For+Me)	140	A

Labeled moiety	Parent ion m/z	Labelling origin	Transition m/z	Monitored ^{13}C trans
Unlabeled	399	--	97	--
A	400	+1 (For)	97	--
C	400	+1 (Me)	97	--
A	401	+2 (2For or Gly)	97	--
A + C	401	+2 (For+Me)	97	--
A	402	+3 (Gly+For)	97	--
A + C	402	+3 (2For or Gly+Me)	97	--
A	403	+4 (Gly+2For)	97	--
A + C	403	+4 (Gly+For +Me)	97	--
B	404	+5 (R5P)	102	B
A + C	404	+5 (Gly+2For +Me)	97	--
A + B	405	+6 (R5P+For)	102	B
B + C	405	+6 (R5P+Me)	102	B
A + B	406	+7 (R5P+Gly or 2For)	102	B
A + B + C	406	+7 (R5P+For +Me)	102	B
A + B + C	407	+8 (R5P+Gly or 2For+Me)	102	B
A + B	407	+8 (R5P+Gly +For)	102	B
A + B	408	+9 (R5P+Gly +2For)	102	B
A + B + C	408	+9 (R5P+Gly +For+Me)	102	B
A + B + C	409	+10 (R5P+Gly +2For+Me)	102	B

Labeled moiety	Parent ion m/z	Labelling origin	Transition m/z	Monitored ^{13}C trans
Unlabeled	399	--	250	--
A	400	+1 (For)	251	A
C	400	+1 (Me)	250	--
A	401	+2 (2For or Gly)	252	A
A + C	401	+2 (For+Me)	251	A
A	402	+3 (Gly+For)	253	A
A + C	402	+3 (2For or Gly+Me)	252	A
A	403	+4 (Gly+2For)	254	A
A + C	403	+4 (Gly+For +Me)	253	A
B	404	+5 (R5P)	255	B
A + C	404	+5 (Gly+2For +Me)	254	A
A + B	405	+6 (R5P+For)	256	A + B
B + C	405	+6 (R5P+Me)	255	B
A + B	406	+7 (R5P+Gly or 2For)	257	A + B
A + B + C	406	+7 (R5P+For +Me)	256	A + B
A + B + C	407	+8 (R5P+Gly or 2For+Me)	257	A + B
A + B	407	+8 (R5P+Gly +For)	258	A + B
A + B	408	+9 (R5P+Gly +2For)	259	A + B
A + B + C	408	+9 (R5P+Gly +For+Me)	258	A + B
A + B + C	409	+10 (R5P+Gly +2For+Me)	259	A + B

Labeled moiety	Parent ion m/z	Labelling origin	Transition m/z	Monitored ^{13}C trans
Unlabeled	399	--	298	--
A	400	+1 (For)	299	A
C	400	+1 (Me)	299	C
A	401	+2 (2For or Gly)	300	A
A + C	401	+2 (For+Me)	300	A + C
A	402	+3 (Gly+For)	301	A
A + C	402	+3 (2For or Gly+Me)	301	A + C
A	403	+4 (Gly+2For)	302	A
A + C	403	+4 (Gly+For +Me)	302	A + C
B	404	+5 (R5P)	303	B
A + C	404	+5 (Gly+2For +Me)	303	A + C
A + B	405	+6 (R5P+For)	304	A + B
B + C	405	+6 (R5P+Me)	304	B + C
A + B	406	+7 (R5P+Gly or 2For)	305	A + B
A + B + C	406	+7 (R5P+For +Me)	305	A + B + C
A + B + C	407	+8 (R5P+Gly or 2For+Me)	306	A + B + C
A + B	407	+8 (R5P+Gly +For)	306	A + B
A + B	408	+9 (R5P+Gly +2For)	307	A + B
A + B + C	408	+9 (R5P+Gly +For+Me)	307	A + B + C
A + B + C	409	+10 (R5P+Gly +2For+Me)	308	A + B + C

For=Formate; Gly=Glycine; Me=Methionine; R5P=Ribose-5'-phosphate

Figure 4. SAM incorporation of labeled carbons. Barplots representing isotopologue abundance for S-adenosylmethionine (SAM) (a), abundance in percentages of carbons incorporated by moiety or combined moieties (A,B,C) in an isotope tracing experiment. Brain cells were incubated for 20h or 72h in labeled C13 glucose or glutamine. Names on the x axis correspond to: precursor ion (e.g. 814); labeled moiety or moieties (e.g. AC); original mass (M) plus the number of extra carbon/neutrons incorporated (e.g. +4). Schematic representation of the fluxes involved in the formation of SAM using labeled glucose or serine (b).

8.3. Concluding remarks

Our first method provides a unique approach to characterize epigenetically relevant metabolites by taking advantage of two complementary targeted metabolomic techniques. To our knowledge, it is the only targeted metabolomics method available to measure formic acid, SCFAs and other key epigenetic metabolites with a single extraction in the same biological sample and without chemical derivatization. Due to avoiding chemical derivatization, we believe that this method offers a great balance between data generation effort and obtained biological insights. However, the main limitation of our method is that it cannot still fully cover the collection of epigenetically relevant metabolites we intended to cover. Some of these metabolites such as homocysteine or folates, require very specific sample preparation to avoid degradation and ensure their stabilization when extracting them from the matrix (*583*, *588*). Our approach can help to discover new molecular mechanisms underlying epigenetic regulation in health and diseases when combined with next generation sequencing and MS based proteomic technologies. It also opens the door to study how microbiota and hosts interact in a multi-kingdom chemical conversation that goes in both directions (*602*). Metabolites produced by the microbiota can regulate epigenetic marks in the host, which in turn, can regulate genes controlling host metabolism and responses to microbiota.

For us, SCFAs were the most important metabolites in our list of epigenetically relevant metabolites. SCFAs are produced during the anaerobic fermentation of complex polysaccharides from the diet in the colon (*89*) – polysaccharides that reach the mammalian colon without being digested by host enzymes. When absorbed, SCFAs enter the bloodstream and can virtually reach any cell in the body and act as inhibitors or activators of acetyl- and methyltransferases (*549*, *603–605*). Moreover, host gut microbiota can influence the abundance of other metabolites that depend on diet such as: choline, methionine or riboflavin (*527*, *606*, *607*). Then, it is plausible to think that microbiota metabolic activity can influence other host endogenous and epigenetically relevant metabolites, such as succinate (*608*) not directly related to microbiota or diet, but of relevance in the epigenetic regulation landscape.

Our second method brings the opportunity of performing targeted isotope tracing experiments without using expensive heavy labeled standards for method optimization. Knowing the biological origin of each atom in a chemical moiety of a given metabolite, together with experimentally testing the MS2 fragmentation patterns of the unlabeled standards and the structural elucidation of the fragments, have allowed us to identify the position of the labeled carbons and the pathways followed by the labeled atoms. To the best of our knowledge, this approach has not been described previously and allows us to measure MS2 isotopologues fragments. This approach increases the information gained with standard

MS1 tracing experiments, by identifying the exact positions where the labeled atoms are deposited after exposing cells to labeled glucose or glutamine.

The two metabolomic tools we present in this manuscript want to foster the design of new basic and clinical experiments that combine different omic technologies to dissect our previously proposed "nutrient-microbiota metabolism-host epigenetic axis" (*22*). For example: the measurement of epigenetically relevant metabolites in type 2 diabetes patients, together with metagenomic and histone PTMs information could be used to design a diet intervention or a probiotic bacterial consortia to "correct" abundances of epigenetically relevant metabolites and change pathogenic associated epigenetic histone marks in diabetes; or tracing carbons from glucose to the acetyl group of acetyl-CoA, in combination with the ones that end up in acetylated histones as shown in (*609*), would demonstrate how labeled carbons from glucose are added to histones in the form of acetyl group; or feeding labeled inulin to mice and observe if acetyl-CoA and histones incorporate labeled acetyl or methyl groups originated from the microbiota fermentation of the labeled inulin ((*610*)).This last proposed approach would give us another strong evidence of how microbiota can influence host epigenetic marks such as has been demonstrated recently by *Lund et al.* (*611*).

8.4. Contributions

The targeted metabolomics method was designed by Oscar Yanes and myself. I performed most of the method experimental optimization with the help and guidance of Sandra Junza. I tested all conditions and extractions described in the methods section for the optimization of the method, and Sandra tested the final biphasic extraction in serum samples and I optimized the final biphasic extraction in liver and cecum samples. I performed all the experimental measures on the samples presented in the results section of the first method, and analyzed and interpreted the data. Oscar, Sandra and myself conceived the idea of measuring the isotopomers without the need of labeled standards for the optimization. Sandra did the isotopomers calculations and measured them for the UDP-N-acetylglucosamine, and I calculated the isotopomers and measured them for the SAM and acetyl-CoA. Silvia Raineri cultured the cells used for the labeling trace analysis of acetyl-CoA. Marta Kovatcheva, under the supervision of Manuel Serrano, has cultured the cells with labeled glucose for measuring the labeled atoms in the SAM molecule. I created and designed most of the figures in the manuscript and wrote the manuscript draft. Jordi Capellades have helped with the analysis of the labeled data and plotted the SAM labeling. And Oscar, Jordi and myself have edited the manuscript up to the current submitted version.

8.5. Material and methods

Standards and metabolomic reagents.
Methanol (MS grade), acetonitrile (ACN, MS grade), pyridine, methoxyamine hydrochloride (MA), formic acid (MS grade), N-methyl-N-(trimethylsilyl)-trifluoroacetamide (TMS) and ammonium acetate were purchased from Thermo Fisher Scientific (Waltham). All pure standards in table 1, 2,3,4,5,6-Pentafluorobenzyl bromide (PFBBr), hexane and acetone were purchased from Sigma-Aldrich (St. Louis). Acetyl-ADP-ribose was purchased from Santa Cruz Biotechnology (Dallas). Medronic acid was purchased from Agilent Technologies (Santa Clara).

Germ-free and conventional female samples
Liver and cecal content were obtained from nine-week-old female germ-free (GF) and conventional (CV) C57BL/J6 mice strain. The GF mice were procured by the breeding unit Anaxem (INRAE, Jouy-en-Josas, France; Anaxem license number: B78-33-6). The CV mice were purchased from Charles River Laboratories (L'Arbresle, France) and kept in the Anaxem facilities. The GF mice were housed in sterile isolators (Getinge, Les Ulis, France) in individual cages. Fresh defecations were used to ensure sterile conditions weekly by microscopic examination and screening cultures. CV mice were housed in the same type of isolators but non-sterile, to ensure the same environment and stress between experimental groups. In each isolator, mice were kept in enriched home cages containing paper towels, wooden sticks and sterile bedding made of wood shavings, and free access to autoclaved tap water and γ-irradiated (45 kGy) standard chow diet (R03; Scientific Animal Food and Engineering, Augy, France). The experimental procedures were performed in accordance with European guidelines for the care and use of laboratory animals and approved by the ethics committee of the INRAE Research Centre at Jouy-en-Josas (approval reference: 17-14). Ten mice females were used for each microbiota status, with a total n=20 mice. All mice were housed individually for measurement of individual food consumption. Body weight and food were weighted once a week from the age of 4 to 9 weeks. And animal room temperature was maintained at 20–24°C with a strict 12-h light/dark cycle (lights open at 7:30 am). Human serum samples were obtained from lab volunteers and processed to separate cells from serum quick after collection.

Cell culture samples in 13C media
For the analysis of acetyl-CoA and SAM: Brain immortalized cell line U-87 MG (ATCC® HTB-14™) (WT) cells were plated in a density of 1 million per T75 plate in 10 mL complete ATCC EMEM 30-2003 medium and have 5 replicates per condition. Two conditions were tested,^{13}C labeled glucose and ^{13}C labeled glutamine. After 8 h, the medium was changed to ^{13}C-Glucose, add 10 mL per T75 (Nacalai Tesque 09848 + ^{13}C Glucose + Sodium Pyruvate + FBS); or ^{13}C-Glutamine, add 10 mL per T75 (Sigma M5650 + ^{13}C Glutamine + Sodium Pyruvate + FBS). A complete description of medium composition can be found in the supplementary *table 4*.

For the analysis of UDP-GlcNAc: WT fibroblasts derived from mouse embryonic stem cells were cultured for 3 days plus 16h in fully labeled ^{13}C-glucose media (13C6-glucose, Sigma-Aldrich). WT cells were cultured in DMEM high glucose (21969035, ThermoFisher) medium supplemented with 10% v/v FBS (SH30066.03, HyClone), 20 mM Hepes, pH 7.3, 100 μM non-essential amino acids (11140035, ThermoFisher), 2 mM L-glutamine (25030-024, ThermoFisher), 100 μM β-mercaptoethanol

(31350010, ThermoFisher) and 100 U/mL Penicillin-Streptomycin (15140122, ThermoFisher), 25 mM glucose, 1 mM sodium pyruvate and 44 mM sodium bicarbonate.

Metabolite extraction methods on mice samples

Frozen lyophilized and pulverized mice liver was used for optimizing and choosing the best metabolite extraction method. Two different types of extractions were tested: 1) monophasic extractions, using one or more miscible solvents and having a homogeneous mixture; and 2) biphasic extractions, using two or more solvents forming two immiscible phases. Once the final extraction method was chosen, we tested the protocol in two additional matrices: mouse caecum content and human serum.

Monophasic extractions

ACN/MeOH/H_2O. A volume of 400 µL of acetonitrile methanol water solution (ACN:MeOH:H_2O in a 4:4:2 volume proportion) was added to 5 mg of pulverized liver tissue and vortex for 1 min. Here, 3 different pH (pH = *3.2*, *6.6* and *9.1*) of the extraction solution were tested. The sample was incubated in liquid nitrogen for 30 s, sonicated for 20-30 s in a water bath and vortexed for 30 s, repeating these steps up to a total of three times. Next, the sample was incubated at -20ºC for 60 min and then centrifuged 10 min at 22000 g (4ºC). Finallys 100 µL of supernatant was transferred into HPLC vial insert.

MeOH/H_2O (8:2 and 1:1). A volume of 500 µL of methanol water (MeOH/ H_2O, in a 8:2 or 1:1 volume proportion) solution was added to 5 mg of pulverized liver tissue, vortexed for 1 min followed by 3 cycles of incubation in liquid nitrogen for 30 s, sonicated for 30 s in a water bath and vortexed for 30 s. The extraction was incubated on ice for 60 min to allow proteins to precipitate and then centrifuged 10 min at 22000 g (4ºC), 50 µL of supernatant was transferred into a HPLC vial insert for LC-MS injection.

Metaphosphoric acid (MPA). A volume of 300 µL of cold acetonitrile water (ACN/H_2O, 1/1) solution with 1 % of metaphosphoric acid was added to 5 mg of pulverized liver tissue, vortexed for 30 s, followed by 3 cycles of incubation in liquid nitrogen for 30 s, sonicated for 30 s in a water bath and vortexed for 30 s. After 2 hours incubation at -20ºC the samples were centrifuged 10 min at 22000 g (4ºC), and 100 µL of supernatant was transferred to a HPLC vial insert.

2-Mercaptoethanol (BME). A volume of 500 µL of 50 mM phosphate buffer (1x), at pH=7, with 1 % of ascorbic acid and 0.1 % 2-mercaptoethanol was added to 5 mg of pulverized liver tissue, vortexed for 1 min, followed by 3 cycles of incubation in liquid nitrogen for 30 s, sonicated for 30 s in a water bath and vortexed for 30 s. After 60 min incubation on ice, the sample was centrifuged 10 min at 22000 g (4ºC), and 100 µL of supernatant transferred to an HPLC vial insert.

Biphasic extractions

H_2O phosphoric 15% solution: Methyl Tert butyl Ether (Phospho/MTBE). A volume of 200 µL of water with 15 % of phosphoric acid was added to 5 mg of pulverized liver tissue and vortex for 1 min, followed by 3 cycles of incubation in liquid nitrogen for 30 s, sonicated for 30 s in a water bath and vortexed for 30 s. Then 200 µL of methyl tert butyl ether (MTBE) was added and vortexed for 1 min. After 60 min incubation on ice the sample was centrifuged 10 min at 22000g (4 ºC) and formed two separated

phases. The acidic H_2O was the bottom phase and MTBE was the top phase. 100 µL of the acidic H_2O phase were diluted in 900 µL of pure acetonitrile (ACN) followed by 1h incubation at -20ºC to facilitate protein precipitation. Then the samples were centrifuged for 10 min at 22000 g (4 ºC) and 100 µL of supernatant (bottom phase) was transferred to an HPLC vial, ready for LC-MS analysis. 100 µL of MTBE (top phase) was directly transferred after the first centrifugation to an HPLC vial insert for direct GC-MS analysis.

H_2O formic 0.1 %/Diethyl Etether (H_2O/DEE). A volume of 400 µL of water with 0.1 % of formic acid was added to 5 mg of pulverized liver tissue and vortex for 1 min, followed by 3 cycles of incubation in liquid nitrogen for 30 s, sonicated for 30 s in a water bath and vortexed for 30 s. Then 400 µL of diethyl ether (DEE) was added and vortexed for 1 min. After 60 min incubation on ice the sample was centrifuged 10 min at 22000 g (4 ºC) and formed two separated phases. Water (H_2O) was the bottom phase and diethyl ether (DEE) was the top phase, from which 400 µL of each phase were separated in different HPLC vials. For LC-MS analysis, an aliquot of 50 µL of aqueous (bottom) phase was transferred to a HPLC vial insert for direct LC-MS analysis. For GC-MS analysis, both phases (H_2O and DEE) were analyzed with the following downstream preparation: H_2O phase was frozen at -80 ºC followed by 2 or more hours in the lyophilizator until dry; and DEE phase was dried under N_2 flux. Lyophilized or dry samples were derivatized as described here. Chemical derivatization. Metabolite standards for method optimization or liver extracts to be analyzed by the GC-MS method were chemically derivatized as follow: 1) adding 40 µL of methoxyamine in pyridine (30 µg/µL) and incubated for 45 minutes at 60°C; and 2) adding 25 µL of N-methyl-N-trimethylsilyltrifluoroacetamide (TMS) with 1% trimethylchlorosilane (Thermo Fisher Scientific) and incubated for 30 minutes at 60°C. The 65 µL of derivatized sample extract or metabolite standards were transferred into HPLC vial inserts for GC-MS analysis.

H_2O formic 0.1 %/Methanol/Diethyl Ether (Mix). This extraction was a mix between the H_2O/DEE extraction and the MeOH/H_2O (1/1) with 0.1 % of formic acid extraction. A volume of 500 µL of MeOH/H_2O (1/1) extraction with 0.1 % of formic acid were added to the sample, followed by the 3 cycles of incubation in liquid nitrogen for 30 s, sonicated for 30 s in a water bath and vortexed for 30 s. Then 500 µL of DEE was added and vortexed for 1 min. After 60 min incubation on ice, the sample was centrifuged 10 min at 22000 g (4 ºC) and formed two separated phases. Water (H_2O) was the bottom phase and diethyl ether (DEE) was the top phase, 50 µL of the water phase were separated for LC-MS analysis. The upper phase (DEE) was separated in a different HPLC vial to be dried under N_2 flow and derivatizated as described in the *H_2O/DEE* extraction.

2,3,4,5,6-Pentafluorobenzyl bromide (PFBBr). A volume of 20 µL of phosphate buffer (0.5M pH 8.0) was added to 50 µL of human serum. Add 130 µL of PFBBr solution (100 mM in acetone) and vortex for 1 min followed by an incubation of 15 min at 60 ºC. After the incubation, wait 2-3 min at RT to cold down the solution and add 330 µL of hexane and vigorously vortex. After vortexing two phases are clearly separated, corresponding the upper phase to the organic solvent containing SCFAs. 100 µL of the supernatants was transferred to a GC vial for analysis.

LC-MS/MS analysis of epigenetically relevant metabolites

O*n standards and test samples*: LC-MS method optimization was performed using around 1ppm dilutions of pure metabolites standards in *table 1*, except for the SCFAs. Pure metabolite standards and sample extracts were analyzed by LC-MS. Metabolites in the sample extractor mixture of pure standards were separated using a 1290 Infinity Series liquid chromatograph (Agilent Technologies), coupled with positive and negative electrospray ionization (-ESI & +ESI) at the same time (*612*), followed by spectral mass data acquisition with a 6490 triple quadrupole QqQ mass spectrometer (Agilent Technologies). The mass spectrometer was operated in multiple reaction monitoring (MRM), and samples for LC-MS analysis were not derivatized. Extracted metabolites or mixtures of pure standards were kept in vials at -80ºC until the LC-MS analysis was performed. Standard mixtures and metabolite extractions were separated using one of the following chromatographic columns: Luna Omega 1.6 µm Polar C18 100 Å, LC Column 100 x 2.1 mm (Phenomenex) as reverse phase column; and ACQUITY UPLC BEH HILIC Column, 130Å, 1.7 µm, 2.1 mm X 150 mm (Waters), InfinityLab Poroshell 120 HILIC-Z (HILIC-Z), 2.1 x 100 mm, 2.7 µm (PEEK lined) (Agilent Technologies) or the SeQuant® ZIC-pHILIC (pHILIC) 5µm polymer 150 x 2.1 mm (Merck) as hydrophilic interactions chromatography (HILIC) columns. The HILIC-Z and SeQuant columns have a zwitterionic stationary phase, having at the same time positive and negative charges, allowing the retention of challenging polar and charged metabolites and with a bigger dynamic range. All columns were operated with their correspondent precolumn.

On experimental samples: Metabolites in liver extracts were separated by using either the HILIC-Z or the pHILIC columns, using either 50 mM ammonium acetate with 5 µM of medronic acid solution or a 20 mM ammonium acetate with 2.5 µM of medronic acid solution, respectively as mobile phase A and 100% acetonitrile (ACN) as phase B. For the HILIC-Z , the mobile phase flux was set to 0.4 mL/min with a linear gradient elution that started at 98% B (time 0-2 min), followed by an isocratic gradient from 98% and finishing at 40 % B (time 2-9 min), then back to 98% B (time 9-9.5 min) and a holding time 3 min and a half at 98% B (time 9.5-13 min) to allow system stabilization. For the pHILIC, the mobile phase flux was set to 0.2 mL/min with a linear gradient elution that started at 85% B (time 0-2 min), followed by an isocratic gradient from 85% and finishing at 40 % B (time 2-12 min), then back to 85% B (time 12-12.5 min) and a holding time 6 min and a half at 85% B (time 12.5-19 min) to allow system stabilization. Ions were generated using positive and negative electrospray ionization (+ESI & -ESI) and spectral data measured with a 6490 QqQ mass spectrometer (Agilent Technologies) operated in both positive and negative ion mode. The injection volume was set to 3 µL. The mass spectrometer parameters were: drying and sheath gas temperatures 270°C and 400°C, respectively; source and sheath gas flows 15 and 11 L/min, respectively; nebulizer flow was set to 35 psi; positive and negative capillary voltage were both set at 3000V; nozzle voltages were 1000V and -1500V, in positive and negative respectively; and iFunnel in positive HRF and LRF 130 and 100V, respectively; and iFunnel in negative HRF and LRF 110 and 60V, respectively. The ions and transitions that have been monitored can be found in supplementary *table 1* in the parent m/z, 1st and 2nd m/z transitions columns, together with each collision energy (CE) used. We manually quantified all pure standards (for method optimization) and metabolite extraction peaks with the Qualitative Analysis of MassHunter Workstation (Agilent Technologies).

GC-MS/MS analysis of SCFAs

O*n standards and test samples*: GC-MS method optimization was performed by using around 1ppm dilutions of pure standards of derivatized and non-derivatized formic, acetic, propionic and butyric acid (SCFAs, in *table 1)*. Pure metabolite standards and sample extracts analyses performed by GC-MS, were separated using a 7890A gas chromatograph (Agilent Technologies), coupled with one of the two different ionization sources tested: electron impact (EI) or chemical ionization (CI), followed by spectral mass data acquisition with a 7000 QqQ mass spectrometer (Agilent Technologies). The mass spectrometer was operated in multiple reaction monitoring (MRM) for derivatized and non-derivatized samples. Derivatized extracts or standards were injected (1µL) into the gas chromatograph with a split inlet and a J&W Scientific DB5−MS+DG (5MS) stationary phase column of 30 m × 0.25 mm i.d., 0.1 µm film (Agilent Technologies). Non-derivatized samples or pure standards analyzed were injected (1µL) into the gas chromatograph system with a split inlet in a J&W Scientific HP-FFAP (FFAP) stationary phase column 30m × 0.25mm i.d., 0.25 µm film (Agilent Technologies).

On samples: The extracted liver SCFAs in the organic phase (top) of the *Phospho/MTBE* extraction were separated by using a J&W Scientific HP-FFAP (FFAP) stationary phase column (30m × 0.25mm i.d., 0.25 µm film, Agilent Technologies) when samples were non-derivatized and into a J&W Scientific DB5−MS+DG (5MS) stationary phase column (30 m × 0.25 mm i.d., 0.1 µm film, Agilent Technologies) when samples were derivatized. Sample metabolites were separated with a 7890A gas chromatograph (Agilent Technologies), coupled with an electron impact (EI) ionization source, followed by spectral mass data acquisition with a 7000 QqQ mass spectrometer (Agilent Technologies). The GC-MS/MS conditions for the derivatized samples were as follow: the 5MS column was used, carrier gas was helium at 7.5 mL/min flow, split ratio was set to 2:1, oven temperature was set to 50 ºC for 2 min, followed by a ramp of 30 ºC/min up to 220 ºC and holding this temperature for 3 min, source heather was set to 300 ºC and ionization was achieved by electron impact (EI) at 70 eV. The GC-MS conditions for the non-derivatized samples were as follow: the FFAP column was used, carrier gas was helium at 54.5.1 mL/min flow, split ratio was set to 10:1, oven temperature was set initially to 40 ºC for 0 min, followed by a ramp of 12 ºC/min up to 130 ºC, then a ramp of 30 ºC/min up to 250 ºC and holding this temperature for 5 min, source heather was set to 250 ºC and ionization was achieved by electron impact (EI) at 70 eV. We manually quantified all metabolite extraction peaks with the Qualitative Analysis of MassHunter Workstation (Agilent Technologies).

13C flux analysis from glucose and glutamine to SAM, acetyl-CoA and UDP-N-acetylglucosamine

Metabolomics analyses were performed on cultured brain cell line U-87 MG with ^{13}C media as follows: at 20h and 72h media was aspirated and washed one time with cold PBS (phosphate buffer or any other physiological buffer) to avoid carry over from media components. In order to extract the metabolites from the cells, 1mL of extraction solution (80:20 methanol:water) prechilled at -80ºC was added to the cells and incubated for 15 min at -80ºC to break the cells. After, the cells were scraped off the dish with a cell scraper and the suspension (cells in the extraction solution) transferred into a clean 1.5 mL, followed by 10 min centrifugation at 2000g at 4ºC to pellet cellular debris. Supernatant was transferred into a new 1.5 mL tube and set aside on ice. The metabolite extraction was repeated two more rounds on the cell pellets, using 125 µL of extraction solution. The corresponding supernatants were combined with the supernatant of the first round of extraction in the plate and stored at -80ºC. Supernatants were

dried under N_2 gas flow until completely dry, in order to resuspend them in the same volume. The dried extracts were resuspended in 400 µL of extraction solution (80:20 methanol:water) and 100 µL were put in a HPLC vial for targeted analysis.
The extracted metabolites were separated using an InfinityLab Poroshell 120 HILIC-Z (Agilent Technologies) column, using a 50 mM ammonium acetate with 5 µM of medronic acid as phase A and 100% acetonitrile (ACN) as phase B. The mobile phase flux was set to 0.4 mL/min with a linear gradient elution that started at 98% B (time 0-2 min), followed by an isocratic gradient from 98% and finishing at 40 % B (time 2-9 min), then back to 98% B (time 9-9.5 min) and a holding time 3 min and a half at 98% B (time 9.5-13 min) to allow system stabilization. Ions were generated using electrospray ionization (ESI) and spectral data measured with a 6490 QqQ mass spectrometer (Agilent Technologies) operated in both positive and negative modes. The injection volume was 5 µL. The mass spectrometer parameters were the same as described in the LC-MS/SM methods section. The parent ions, transitions and collision energies that have been monitored can be found in supplementary table 3. We calculated most relevant and fully labeled isotopologues for SAM, acetyl-CoA and UDP-Glc-NAc based on: the total number of carbon atoms possibly labeled; pathways involved in the synthesis of their moieties from labeled glucose or glutamine; and previous knowledge of fragmentation patterns generated by MS analysis of the unlabelled metabolites. All samples were manually quantified by extracting and integrating the MRM peak intensities with Qualitative Analysis of MassHunter Workstation (Agilent Technologies).

A previous biological knowledge about the origin of the moieties forming the three metabolites tested in the ^{13}C labeled flux experiment was required. The three metabolites (SAM, acetyl-CoA and UDP-Glc-NAc) have a ribose moiety where all 5 carbons are susceptible to become labeled when coming from the pentose phosphate pathway, in both, labeled glucose and glutamine. The acetyl group from acetyl-CoA and UDP-Glc-NAc can be labeled from the oxidative carboxylation via oxaloacetate and pyruvate when glucose or glutamine are labeled, but only can be labeled from the reductive carboxylation pathway via alpha-ketoglutarate and citrate, when glutamine was labeled but not glucose. Carbon 4 and 5 in the purine nucleotide adenine present in SAM and acetyl-CoA, can be labeled via glycolysis (from labeled glucose) or gluconeogenesis (from labeled glutamine) through the amino acid synthesis pathway. In SAM, the carbon atom in the methyl group covalently bound to the sulfur atom can be labeled. Only if, the ^{13}C is originated from the amino acid serine in the oxidative carboxylation of glutamine or glycolysis from glucose, and through the amino acid synthesis pathway first, and one carbon metabolism later. In UDP-Glc-NAc, the 6 carbon atoms from glucosamine can be labeled from glucose via glycolysis and hexose phosphate pathway, and from glutamine via the oxidative carboxylation first and gluconeogenesis later. Carbon atoms 4, 5 and 6 in the pyrimidine nucleotide uracil in UDP-Glc-NAc, can be labeled from aspartate through the amino acid synthesis pathway, via glycolysis when starting from glucose or, via oxidative carboxylation when starting from glutamine.

Another important step to accurately generate trustable isotope labeling data is to subtract the natural isotopic distribution of previous isotopologues or parent ion in each of the following isotopologues. To measure a more accurate abundance of a parent ion called M+0, and its isotopologues named M+1, M+2…M+n in a glucose carbon ^{13}C labeling experiment, the natural isotopic distribution of M+0 need to be taken into account when measuring M+1, M+2…, because will influence the abundance of the ions M+1, M+2…M+n of the same molecule. The natural isotopic distribution of the M+1 (with the influence of the isotope distribution of M+0 already subtracted) will also influence the abundance of

the ions M+2, M+3…M+n, and the same for M+2 natural isotopic distribution influencing M+3…M+n and so on. To correct for the contribution of naturally occurring isotope distributions among the isotopologues that are monitored, the natural isotope distribution of the previous parent or isotopologue need to be subtracted. To do so, the theoretical natural isotope distribution for each of the m/z monitored (parent ion or previous isotopologue) were obtained using the enviPat web at https://www.envipat.eawag.ch/. From which the highest pattern profile was obtained for each parent (or isotopologue) using the molecular formula and taking into account the number of carbon ^{13}C atoms present. The influence isotopic threshold was set to $1x10^{-10}$ in order to account for most of the natural distribution of ^{13}C atoms. Once this matrix with contributions in the isotopic distributions of all monitored m/z values for the same formula were obtained, the following operations were performed up until the last isotopologue species:

$(M+0)_r = (M+0)_m$ => the M+0 has no natural isotopic distribution influence from a predecessor isotopologue with the same exact chemical formula

$$(M+1)_r = (M+1)_m - ((M+0)_m * (M+0)_1 / 100)$$

$$(M+2)_i = (M+2)_m - ((M+0)_m * (M+0)_2 / 100)$$

$$(M+2)_r = (M+2)_i - ((M+1)_r * (M+1)_1 / 100)$$

$$(M+3)_i = (M+3)_m - ((M+0)_m * (M+0)_3 / 100)$$

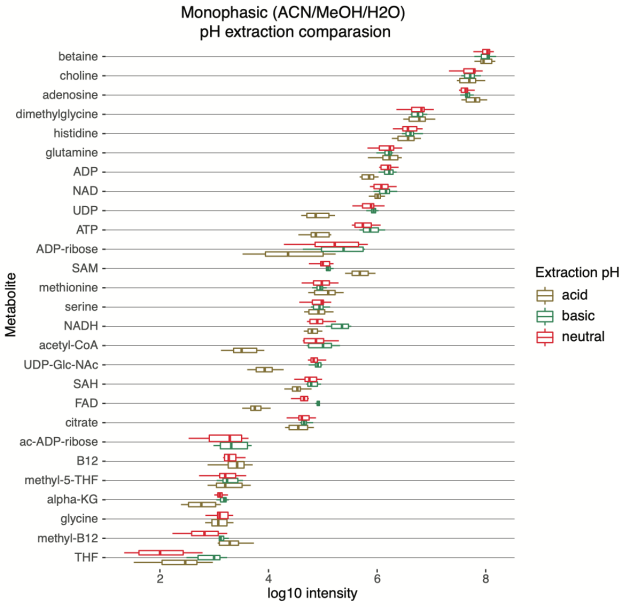
$$(M+3)_{ii} = (M+3)_i - ((M+1)_r * (M+1)_2 / 100)$$

$$(M+3)_r = (M+3)_{ii} - ((M+2)_r * (M+1)_1 / 100)$$

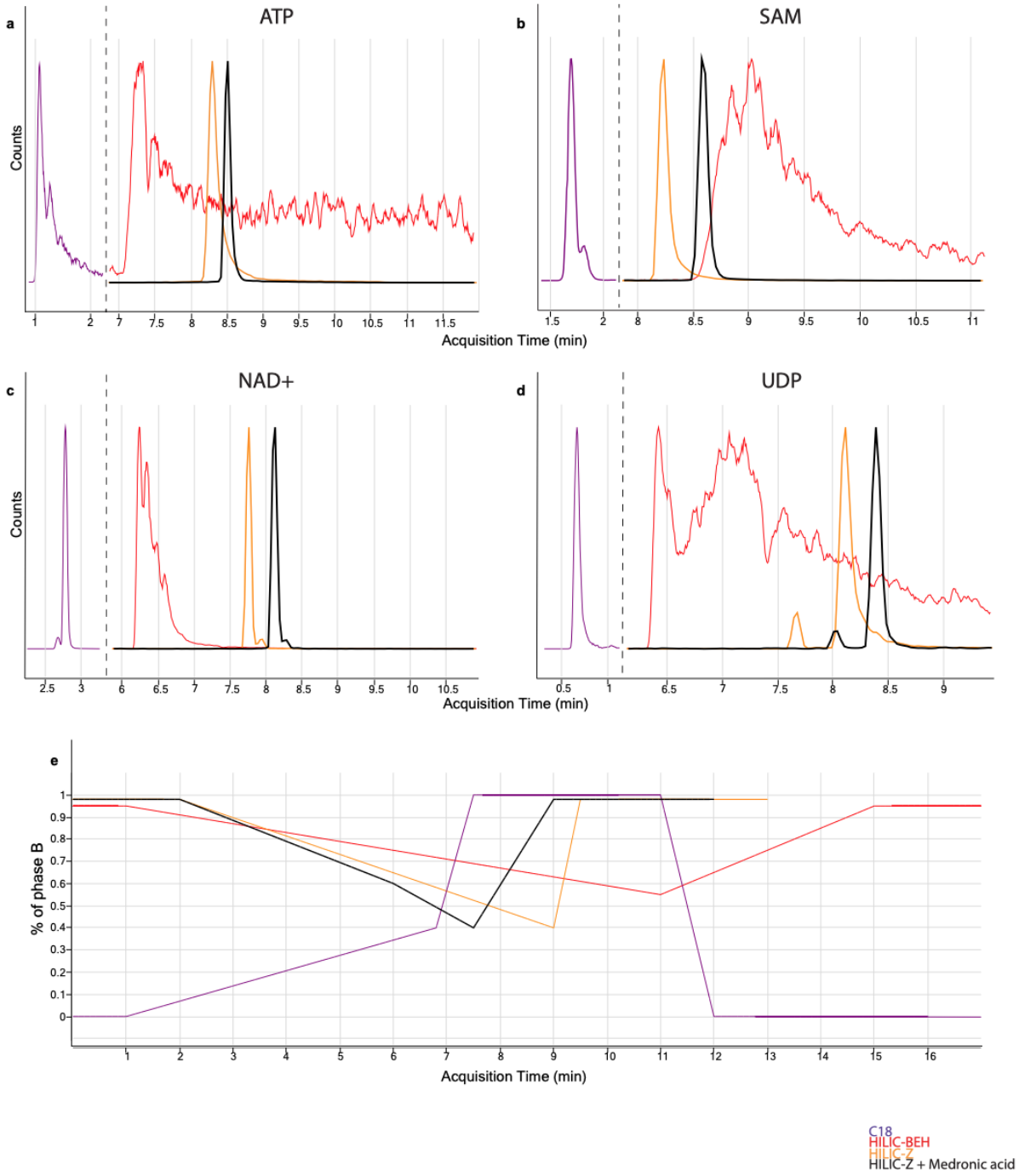
where M + (isotopologue number)$_m$ represents the measured abundance for the M + (isotopologue number), M + (isotopologue number)$_{i \text{ or } ii \text{ or } iii...}$ represents the intermediate abundances for the M + (isotopologue number), M + (isotopologue number)$_r$ represents the real abundances for that particular M + (isotopologue number) after subtracting the corresponding influences of predecessor isotopologues or parent ion, and where M + (isotopologue number)$_{1 \text{ or } 2 \text{ or } 3...}$ represents the theoretical abundances of the isotopologues for that particular specie calculated with enviPat expressed in percentage.

SUPPLEMENTARY material of Chapter 2

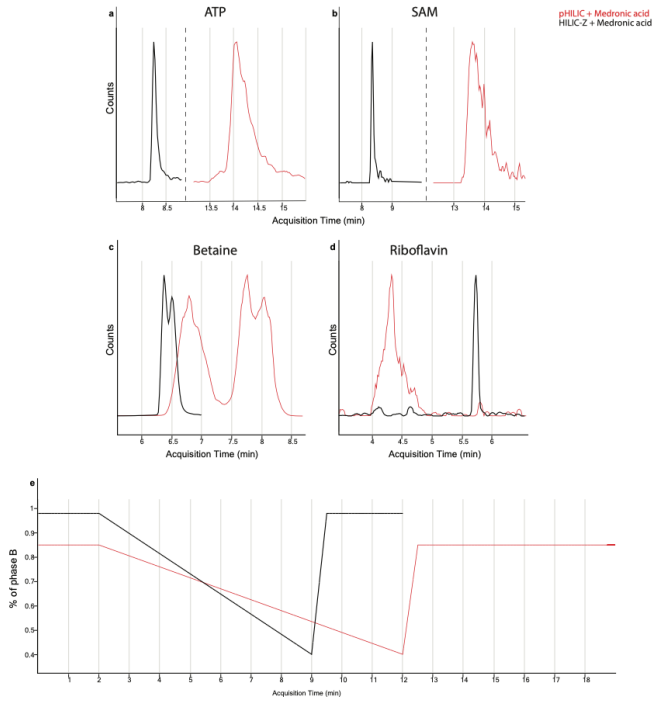
Supplementary figures:



Supplementary figure 1. Influence of pH on metabolite extraction.
Liver log10 intensities of metabolites extracted with the acetonitrile:methanol:water (4:4:2, v:v:v) at pH=3.2 (acid), pH=6.6 (neutral) and pH=9.1 (basic) adjusted extraction solutions.



Supplementary figure 2. Positively charged metabolites or with a phosphate group dictate LC conditions**.** (a-d) Multiple reaction monitoring (MRM) extracted chromatograms of ATP, SAM, NAD+ and UDP compared across four different chromatographic conditions. (e) Chromatographic elution profile showing the percentage of B (organic) phase. C18: reverse phase C18 column; HILIC-BEH: hydrophilic interactions liquid chromatographic (HILIC) ethylene bridged hybrid (BEH); HILIC-Z: HILIC zwitterionic column; HILIC-Z + medronic: HILIC-Z column run with medronic acid in the aqueous phase.



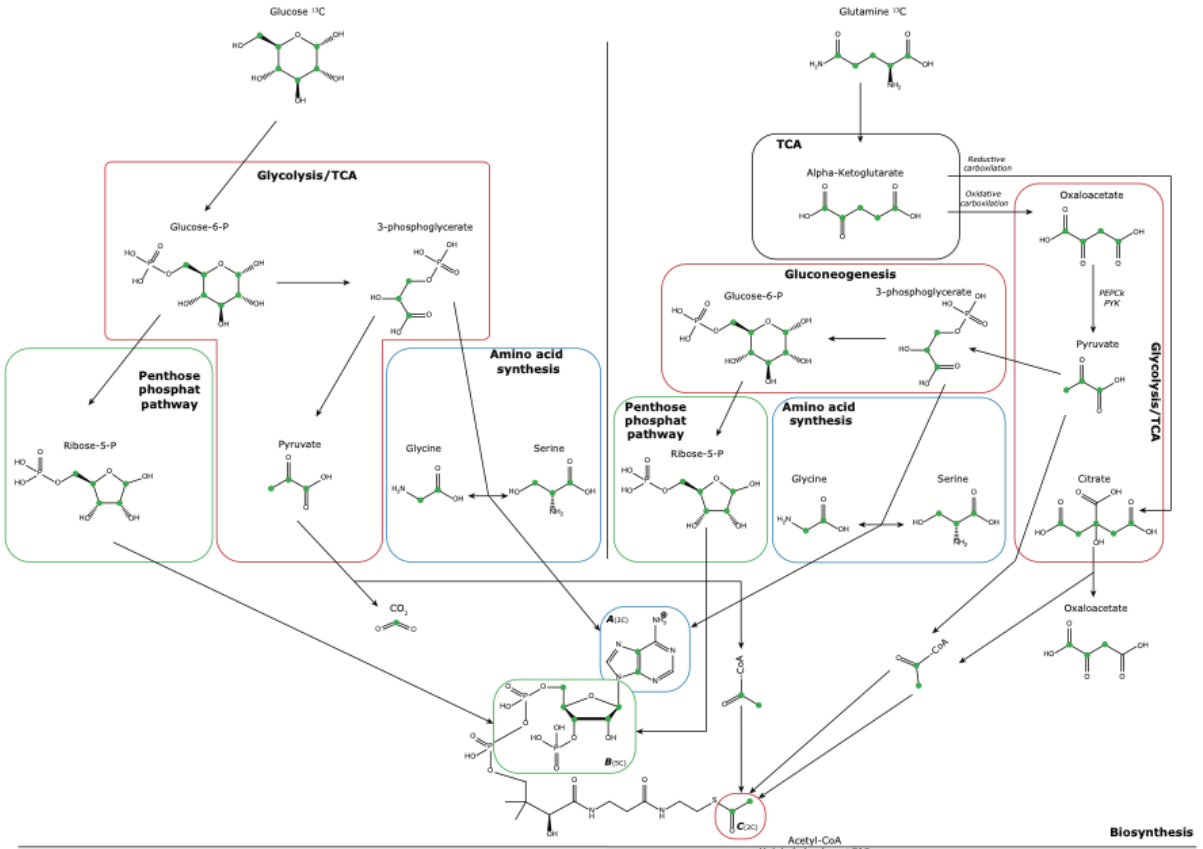
Supplementary figure 3. Zwitterionic column comparison.
(a-d) Multiple reaction monitoring (MRM) extracted chromatograms of ATP, SAM, betaine and riboflavin compared using two different zwitterionic columns. (e) Chromatographic elution profiles showing the percentage of B (organic) phase. pHILIC: SeQuant® ZIC-pHILIC column (Merck); HILIC-Z: InfinityLab Poroshell 120 HILIC-Z (Agilent).

Supplementary text 1: acetyl-CoA labeling further details

Suppose that the most intense fragment of acetyl-CoA for your instrument setup is the m/z_2=303, with two possible labeled carbon atoms (Supplementary figure 4). These two carbons are originated from glucose: via glycolysis and through pyruvate; or from glutamine: via oxidative carboxylation through the oxaloacetate-pyruvate pathway, or via reductive carboxylation through citrate, both converting glutamine to alpha-ketoglutaric acid initially. Supposing that only these two labeled carbons were incorporated in the whole acetyl-CoA molecule, we would need to add +2 to the precursor ion m/z_0=810 becoming m/z_{0*}=812 and the fragment m/z_2=303 becoming m/z_{2*}=305. But there are at least three moieties in acetyl-CoA that can incorporate labeled carbon atoms from glucose or glutamine (Supplementary figure 4). We take the assumption that all possibly labeled carbon atoms are actually labeled and that there are no intermediate or half labeled moieties. For calculating all labeled possibilities, we assume that fully labeled moieties are independent of each other. We can have labeling in all carbons of all moieties, or not having any labeling in any moiety and all the combinations in between (Tables in supplementary figures 4-6).

b

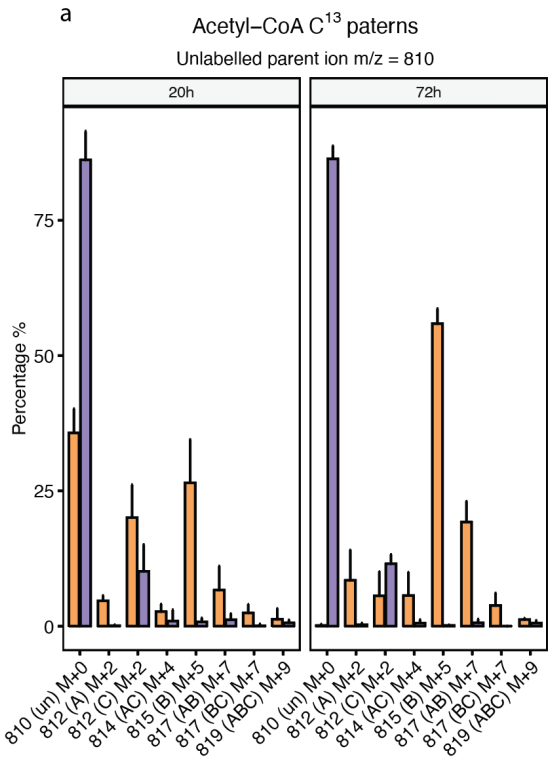
Acetyl-CoA moiety biosinthesys, labelling paterns and LC-QQQ analysis



Labelling	Parent ion m/z	Labeled moiety	Transition m/z	Monitored 13C moiety
Unlabeled	810	na	136	na
Labeled	812	A	138	A
Labeled	815	B	136	na
Labeled	812	C	136	na
Labeled	817	A + B	138	A
Labeled	814	A + C	138	A
Labeled	817	B + C	136	na
Labeled	819	A + B + C	138	A

Labelling	Parent ion m/z	Labeled moiety	Transition m/z	Monitored 13C moiety
Unlabeled	810	na	428	na
Labeled	812	A	430	A
Labeled	815	B	433	B
Labeled	812	C	428	na
Labeled	817	A + B	435	A + B
Labeled	814	A + C	430	A
Labeled	817	B + C	433	B
Labeled	819	A + B + C	435	A + B

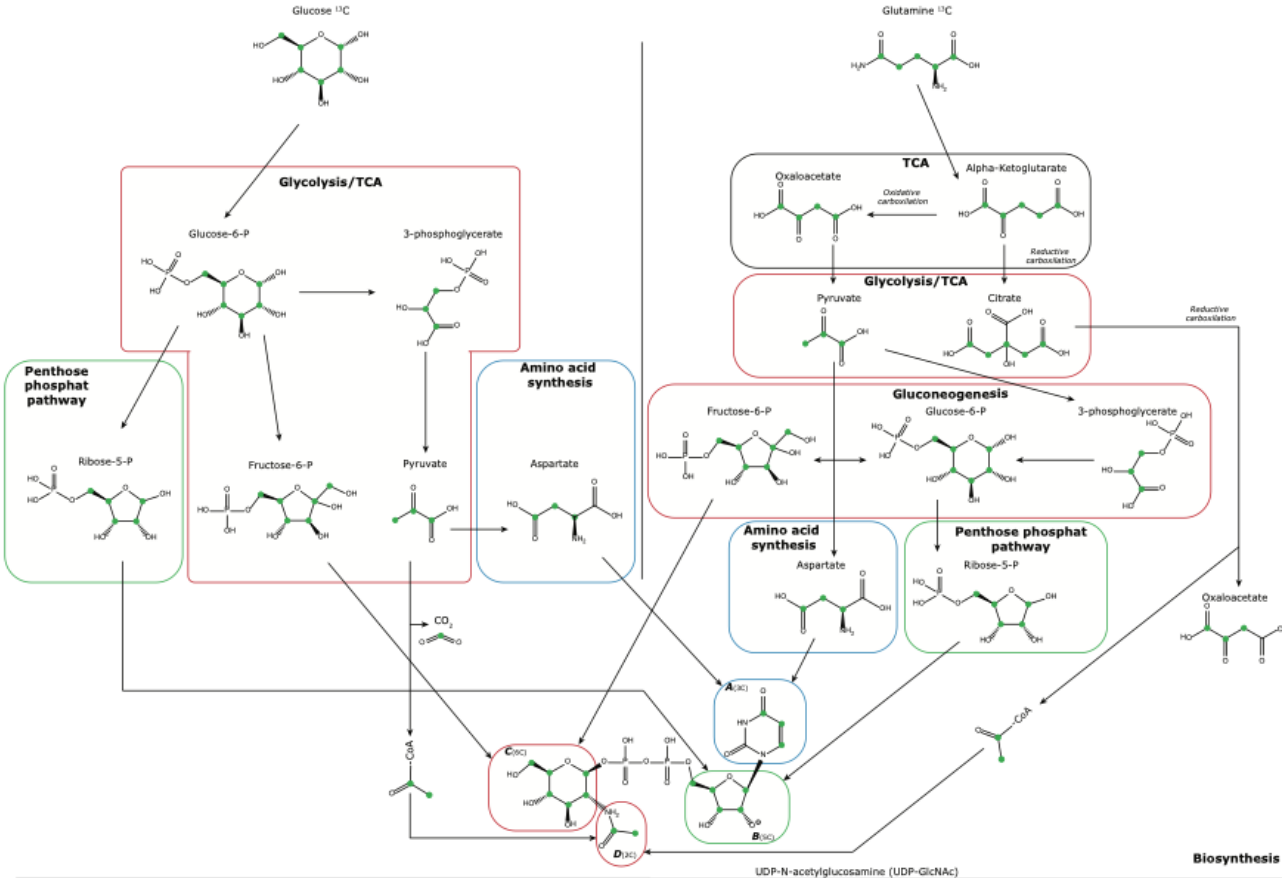
Labelling	Parent ion m/z	Labeled moiety	Transition m/z	Monitored 13C moiety
Unlabeled	810	na	303	na
Labeled	812	A	303	na
Labeled	815	B	303	na
Labeled	812	C	305	C
Labeled	817	A + B	303	na
Labeled	814	A + C	305	C
Labeled	817	B + C	305	C
Labeled	819	A + B + C	305	C



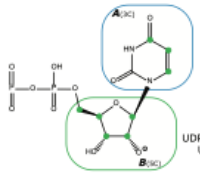
Supplementary figure 4. Acetyl CoA labeling patterns. Barplots showing the enrichment of labeled moieties being enriched after exposure to labeled glucose (orange) or glutamine (purple) for 20h (left panel) or 76h (right panel) (a).
Acetyl-coenzyme A (acetyl-CoA) moieties biosynthesis, fragmentation patterns and calculated isotopologues from labeled glucose or glutamine (b). Biosynthesis and pathways followed by labeled carbon atoms from glucose or glutamine to acetyl-CoA (top). Most intense fragments in the mass spectrometer instrument and calculated moieties isotopologues (bottom).

a

UDP-N-Acetyl-Glucosamine moiety biosinthesys, labelling paterns and LC-QQQ analysis



Mass spectrometry



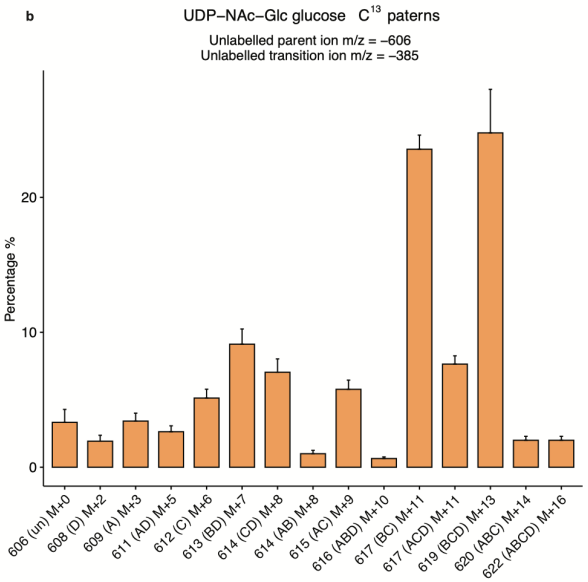
UDP-N-acetylglucosamine (UDP-GlcNAc)
Unlabeled m/z = -606
$C_{17}H_{26}N_3O_{17}P_2^-$

UDP-GlcNAc fragment A + B
Unlabeled m/z = -385
$C_8H_{13}N_2O_{11}P_2^-$

Labelling	Parent ion m/z	Labeled moiety	Transition m/z	Monitored ^{13}C moiety
Unlabeled	-606	na	-385	na
Labeled	-609	A	-388	A
Labeled	-611	B	-390	B
Labeled	-612	C	-385	na
Labeled	-608	D	-385	na
Labeled	-614	A + B	-393	A + B
Labeled	-615	A + C	-388	A
Labeled	-611	A + D	-388	A
Labeled	-617	B + C	-390	B
Labeled	-613	B + D	-390	B
Labeled	-614	C + D	-385	na
Labeled	-620	A+B+C	-393	A + B
Labeled	-616	A+B+D	-393	A + B
Labeled	-617	A+C+D	-388	A
Labeled	-619	B+C+D	-390	B
Labeled	-622	A+B+C+D	-393	A + B

UDP-GlcNAc fragment (na)
Unlabeled m/z = -79
O_3P^-

Labelling	Parent ion m/z	Labeled moiety	Transition m/z	Monitored ^{13}C moiety
Unlabeled	-606	na	-79	na
Labeled	-609	A	-79	na
Labeled	-611	B	-79	na
Labeled	-612	C	-79	na
Labeled	-608	D	-79	na
Labeled	-614	A + B	-79	na
Labeled	-615	A + C	-79	na
Labeled	-611	A + D	-79	na
Labeled	-617	B + C	-79	na
Labeled	-613	B + D	-79	na
Labeled	-614	C + D	-79	na
Labeled	-620	A+B+C	-79	na
Labeled	-616	A+B+D	-79	na
Labeled	-617	A+C+D	-79	na
Labeled	-619	B+C+D	-79	na
Labeled	-622	A+B+C+D	-79	na



Supplementary figure 5. UDP-N-acetylglucosamine labeling patterns. UDP-N-acetylglucosamine (UDP-GlcNAc) moieties biosynthesis, fragmentation patterns and calculated isotopologues from labeled glucose or glutamine.

(a) Biosynthesis and pathways followed by labeled carbon atoms from glucose or glutamine to UDP-GlcNAc (top). Most intense fragments in the mass spectrometer instrument and calculated isotopologues moieties (bottom). (b) Isotopologues of UDP-GlcNAc and percentages of carbons incorporated by moiety or combined moieties in an isotope tracing experiment. WT fibroblasts derived from mouse embryonic stem cells were incubated for 72h in labeled C13 glucose. Names on the x axis correspond to: precursor ion (e.g. 611); labeled moiety or moieties (e.g. AD); original mass (M) plus the number of extra carbon/neutrons incorporated (e.g. +5).

Supplementary tables

Supplementary table 1. Epigenetically relevant metabolites information. Metabolite name, chemical formula, monoisotopic mass, 1st transition (collision energy), 2nd transition (collision energy), extraction with best peak shape, polarity detected, chromatography used, origin for the host, reported epigenetic functions and their references.

Metabolite	Chemical formula	Monoisotopic mass (g/mol)	Parent m/z	m/z 1st transition (CE)	m/z 2nd transition (CE)	Type of extraction with better peak shape	Polarity mode of detection	Metabolomics analytical setup	Origin (Highest contribution for the host)	Epigenetic Function (at least one)	Reference
Formic acid	CH_2O_2	46.005479	46	45()	29()	Biphasic	Pos	GC-QqQ	Microbiome	Methyl donor	[1]
Acetic acid	$C_2H_4O_2$	60.021129	60	45()	43()	Biphasic	Pos	GC-QqQ	Microbiome	Acetyl donor	[2]
Propionic acid	$C_3H_6O_2$	74.036779	74	73()	55()	Biphasic	Pos	GC-QqQ	Microbiome	Acetyl donor and HDAC3 inhibitor	[3,4,5]
Butyric acid	$C_4H_8O_2$	88.052429	88	60()	73()	Biphasic	Pos	GC-QqQ	Microbiome	Acetyl donor and HDACs inhibitor	[6,7,8,9,10]
Glycine	$C_2H_5NO_2$	75.032028	76	30 (8)	48 (4)	Monophasic	Pos	LC-QqQ	Host metabolism	Methyl donor	[11]
Choline	$C_5H_{14}NO^+$	104.107539	104	60 (20)	45 (15)	Both	Pos	LC-QqQ	Diet	Methyl donor	[12]
N, N-Dimethylglycine	$C_4H_9NO_2$	103.063329	104	58 (12)	42 (35)	Both	Pos	LC-QqQ	Host metabolism	Methylation transfer product	[13]
Serine	$C_3H_7NO_3$	105.042593	106	60 (12)	42 (20)	Both	Pos	LC-QqQ	Host metabolism	Methyl donor	[11]
Fumaric acid	$C_4H_4O_4$	116.010959	-115	-71 (4)	-27 (8)	Both	Neg	LC-QqQ	Host metabolism	Inhibitor of TET and JmjC demethylation enzymes	[14]
Succinic acid	$C_4H_6O_4$	118.02609	-117	-73 (8)	-99 (8)	Both	Neg	LC-QqQ	Host and microbiome	Inhibitor of TET and JmjC demethylation enzymes	[15,16,17]
Betaine	$C_5H_{11}NO_2$	117.078979	118	58 (28)	59 (16)	Both	Pos	LC-QqQ	Diet	Methyl donor	[18]
Homocysteine	$C_4H_9NO_2S$	135.0354	136	90 (8)	56 (16)	none	Pos	LC-QqQ	Host metabolism	Methyl acceptor	[13]
Alpha-Ketoglutaric acid	$C_5H_4O_5^{-2}$	144.005873	-145	-101 (4)	-57 (8)	Both	Neg	LC-QqQ	Host metabolism	JmjC co-substrate	[19,20,21]
Glutamine	$C_5H_{10}N_2O_3$	146.069142	147	84 (20)	130 (8)	Both	Pos	LC-QqQ	Host metabolism and diet	Alpha-Ketoglutaric acid precursor	[22]
Methionine	$C_5H_{11}NO_2S$	149.05105	150	56 (12)	104 (6)	Both	Pos	LC-QqQ	Diet	Methyl donor	[23]
Histidine	$C_6H_9N_3O_2$	155.069477	156	110 (12)	83 (28)	Both	Pos	LC-QqQ	Diet	THF precursor	[24]
Glucose	$C_6H_{12}O_6$	180.063388	181	99 (12)	140 (4)	Biphasic	Pos	LC-QqQ	Diet and host metabolism	TET enzyme modulator	[25]
Citrate	$C_6H_5O_7^{-3}$	189.003527	-191	-67 (28)	-57 (20)	Both	Neg	LC-QqQ	Host metabolism	Acetyl donor	[26]
Pantothenic acid (Vitamin B5)	$C_9H_{17}NO_5$	219.110673	220	90 (12)	202 (12)	Both	Pos	LC-QqQ	Diet and microbiome	Acetyl donor	[27]
N-acetylglucosamine	$C_8H_{15}NO_6$	221.089937	222	138 (12)	204 (4)	Biphasic	Pos	LC-QqQ	Diet?	PRC2 substrate and GlcNAc unit	[28,29]
Adenosine	$C_{10}H_{13}N_5O_4$	267.096754	268	136 (16)	119 (48)	Both	Pos	LC-QqQ	Host metabolism and microbiome?	DNMT inhibitor	[30]
Adenosine 5'-monophosphate (AMP)	$C_{10}H_{14}N_5O_7P$	347.063085	348	136 (15)	97 (30)	Both	Pos	LC-QqQ	Host metabolism	Phosphor donor	[31]
Riboflavin (Vitamin B2)	$C_{17}H_{20}N_4O_6$	376.138284	377	243 (20)	172 (35)	Both	Pos	LC-QqQ	Diet and microbiome	FAD precursor and LSD	[27,32]
S-adenosylhomocysteine (SAH)	$C_{14}H_{20}N_6O_5S$	384.121589	385	136 (24)	250 (4)	Both	Pos	LC-QqQ	Host metabolism and microbiome?	Methylation transfer product and methyltransferases inhibitor	[32]
S-adenosylmethionine (SAM)	$C_{15}H_{23}N_6O_5S^+$	399.145064	399	250 (12)	97 (32)	Monophasic	Pos	LC-QqQ	Host metabolism and microbiome?	Methyl donor and co-factor	[32]
Uridine 5'-diphosphate (UDP)	$C_9H_{14}N_2O_{12}P_2$	404.002198	405	97 (24)	113 (36)	Monophasic	Pos	LC-QqQ	Host metabolism	GlcNAc carrier	[33]

Metabolite	Chemical formula	Monoisotopic mass (g/mol)	Parent m/z	m/z 1st transition (CE)	m/z 2nd transition (CE)	Type of extraction with better peak shape	Polarity mode of detection	Metabolomics analytical setup	Origin (Highest contribution for the host)	Epigenetic Function (at least one)	Reference
Adenosine 5'-diphosphate (ADP)	$C_{10}H_{15}N_5O_{10}P_2$	427.029416	428	136 (32)	348 (16)	Monophasic	Pos	LC-QqQ	Host metabolism	Ribosylation carrier	[34]
Dihydrofolic acid (DHF)	$C_{19}H_{21}N_7O_6$	443.155331	-442	-176 (28)	-265 (16)	Monophasic	Neg	LC-QqQ	Microbiome and host metabolism	Methyl donor	[24]
Folic acid (Vitamin B9)	$C_{19}H_{19}N_7O_6$	441.139681	442	295 (12)	120 (40)	Biphasic	Pos	LC-QqQ	Microbiome and host metabolism	Methyl donor	[27,24,35]
Tetrahydrofolic acid (THF)	$C_{19}H_{23}N_7O_6$	445.170981	446	299 (20)	120 (44)	Monophasic	Pos	LC-QqQ	Host metabolism and microbiome	Methyl donor	[24]
N5-Methyl-tetrahydrofolic acid	$C_{20}H_{25}N_7O_6$	459.186632	460	313 (24)	180 (44)	Monophasic	Pos	LC-QqQ	Host metabolism and microbiome	Methyl donor	[24]
Adenosine 5'-triphosphate (ATP)	$C_{10}H_{16}N_5O_{13}P_3$	506.995747	508	136 (44)	97 (40)	Both	Pos	LC-QqQ	Host metabolism	Phosphate donor and co-factor	[34]
ADP Ribose (ADPR)	$C_{15}H_{23}N_5O_{14}P_2$	559.071674	560	136 (36)	348 (16)	Monophasic	Pos	LC-QqQ	Host metabolism	Ribose donor and acetyl group acceptor	[34]
O-Acetyl-ADP-Ribose	$C_{17}H_{25}N_5O_{15}P_2$	601.082239	602	136 (36)	348 (20)	Both	Pos	LC-QqQ	Host metabolism	Acetyl donor	[34]
UDP-N-acetylglucosamine	$C_{17}H_{27}N_3O_{17}P_2$	607.08157	608	204 (8)	138 (48)	Monophasic	Pos	LC-QqQ	Host metabolism	GlcNAc donor	[33]
Nicotinamide adenine dinucleotide (NAD)	$C_{21}H_{28}N_7O_{14}P_2^-$	664.116948	664	136 (60)	428 (24)	Monophasic	Pos	LC-QqQ	Host metabolism	Deacetylases and Sirtuins cofactor	[36,37]
NADH	$C_{21}H_{29}N_7O_{14}P_2$	665.124773	666	136 (40)	137 (52)	Monophasic	Pos	LC-QqQ	Host metabolism	Reduced NAD form	[36,37]
Coenzyme A (CoA)	$C_{21}H_{36}N_7O_{16}P_3S$	767.11521	768	261 (30)	428 (25)	Both	Pos	LC-QqQ		Acetyl group carrier, pantothenic acid precursor and coenzyme	[27]
Flavin adenine dinucleotide (FAD)	$C_{27}H_{33}N_9O_{15}P_2$	785.157135	786	348 (24)	136 (48)	Monophasic	Pos	LC-QqQ	Diet	LSD demethylases cofactor	[32]
Acetyl-Coenzyme A (ac-CoA)	$C_{23}H_{38}N_7O_{17}P_3S$	809.125775	810	303 (36)	136 (60)	Monophasic	Pos	LC-QqQ	Host metabolism	Acetyl donor and acetyltransferase cofactor	[36,38]
Malonyl-CoA	$C_{24}H_{38}N_7O_{19}P_3S$	853.115604	854	303 (32)	347 (48)	Biphasic	Pos	LC-QqQ	Host metabolism	Malonyl donor	[39,40]
Cyanocobalamin (Vitamin B12)	$C_{63}H_{88}CoN_{14}O_{14}P$	1355.575224	678	147 (44)	912 (40)	Biphasic	Pos	LC-QqQ	Diet and microbiome	Methyltransferase cofactor	[27,41]

Supplementary table 2. Extraction protocol and matrix dependent metabolite extraction. Type of extraction with better peak shape for each *epig-metabolite*, depending on the chromatographic method used and the extraction where the metabolites were extracted in each matrix. In this table, both extractions were run under the same LC-MS/MS conditions. Mo: monophasic extraction; Bi: biphasic extraction; empty: not extracted or detected in samples but detected in a mix of pure standards

Metabolite	Type of extraction with better peak shape	Metabolomics analytical setup	Liver	Serum	Caecum
Formic acid	Biphasic	GC-QqQ	Bi	Bi	Bi
Acetic acid	Biphasic	GC-QqQ	Bi	Bi	Bi
Propionic acid	Biphasic	GC-QqQ	Bi	Bi	Bi
Butyric acid	Biphasic	GC-QqQ	Bi	Bi	Bi
Glycine	Monophasic	LC-QqQ			Mo
Choline	Both	LC-QqQ	Mo/Bi	Mo/Bi	Mo/Bi
N, N-Dimethylglycine	Both	LC-QqQ	Mo/Bi	Mo/Bi	Mo/Bi
Serine	Both	LC-QqQ	Mo/Bi	Mo/Bi	Mo/Bi
Fumaric acid	Both	LC-QqQ	Mo/Bi	Mo	Mo
Succinic acid	Both	LC-QqQ	Mo/Bi	Mo/Bi	Mo/Bi
Betaine	Both	LC-QqQ	Mo/Bi	Mo/Bi	Mo/Bi
Homocysteine	none	LC-QqQ			
Alpha-Ketoglutaric acid	Both	LC-QqQ	Mo/Bi	Mo/Bi	Mo/Bi
Glutamine	Both	LC-QqQ	Mo/Bi	Mo/Bi	Mo/Bi
Methionine	Both	LC-QqQ	Mo/Bi	Mo/Bi	Mo/Bi
Histidine	Both	LC-QqQ	Mo/Bi	Mo	Mo/Bi
Glucose	Biphasic	LC-QqQ	Bi	Bi	Bi
Citrate	Both	LC-QqQ	Mo	Mo/Bi	Mo/Bi
Pantothenic acid (vitamin B5)	Both	LC-QqQ	Mo/Bi	Mo/Bi	Mo/Bi
N-acetylglucosamine	Biphasic	LC-QqQ	Bi	Bi	Mo/Bi
Adenosine	Both	LC-QqQ	Mo/Bi	Mo/Bi	Mo/Bi
Adenosine 5'-monophosphate (AMP)	Both	LC-QqQ	Mo/Bi	Mo	Mo/Bi
Riboflavin (vitamin B2)	Both	LC-QqQ	Mo/Bi	Mo/Bi	Mo/Bi
S-adenosylhomocysteine (SAH)	Both	LC-QqQ	Mo/Bi		
S-adenosylmethionine (SAM)	Monophasic	LC-QqQ	Mo/Bi	Mo	Mo
Uridine 5'-diphosphate (UDP)	Monophasic	LC-QqQ	Mo/Bi		Mo
Adenosine 5'-diphosphate (ADP)	Monophasic	LC-QqQ	Mo/Bi	Mo	Mo
Dihydrofolic acid (DHF)	Monophasic	LC-QqQ			Mo
Folic acid (Vitamin B9)	Biphasic	LC-QqQ		Bi	
Tetrahydrofolic acid (THF)	Monophasic	LC-QqQ			Mo
N5-Methyl-tetrahydrofolic acid	Monophasic	LC-QqQ	Mo	Mo	
Adenosine 5'-triphosphate (ATP)	Both	LC-QqQ	Mo/Bi		
ADP Ribose (ADPR)	Monophasic	LC-QqQ	Mo/Bi	Mo	Mo
O-Acetyl-ADP-Ribose	Monophasic	LC-QqQ	Mo	Bi	Mo
UDP-N-acetylglucosamine	Monophasic	LC-QqQ	Mo/Bi	Mo	
Nicotinamide adenine dinucleotide (NAD)	Monophasic	LC-QqQ	Mo/Bi	Mo	Mo
NADH	Monophasic	LC-QqQ	Mo/Bi		Mo
Coenzyme A (CoA)	Biphasic	LC-QqQ	Mo/Bi	Bi	
Flavin adenine dinucleotide (FAD)	Monophasic	LC-QqQ	Mo/Bi		Mo
Acetyl-Coenzyme A (ac-CoA)	Monophasic	LC-QqQ	Mo/Bi		Mo
Malonyl-CoA	Biphasic	LC-QqQ	Bi		
Cyanocobalamin (Vitamin B12)	Both	LC-QqQ		Mo/Bi	

Supplementary table 3. Labeling experiment parent ions and transitions. Parent ions, transitions and collision energies for labeled SAM, Acetyl-CoA and UDP-Gln-NAc moieties.

Metabolite	Labeled moieties	Parent m/z	Transition m/z & (collision energy)
Acetyl-CoA	Na	810	136 (60)
	Na	810	303 (36)
	Na	810	428 (25)
	C	812	136 (60)
	A	812	138 (60)
	A	812	303 (36)
	C	812	305 (36)
	C	812	428 (25)
	A	812	430 (25)
	A + C	814	138 (60)
	A + C	814	305 (36)
	A + C	814	430 (25)
	B	815	136 (60)
	B	815	303 (36)
	B	815	433 (25)
	B + C	817	136 (60)
	A + B	817	138 (60)
	A + B	817	303 (36)
	B + C	817	305 (36)
	B + C	817	433 (25)
	A + B	817	435 (25)
	A + B + C	819	138 (60)
	A + B + C	819	305 (36)
	A + B + C	819	435 (25)

SAM	Na	399	97 (32)
	Na	399	250 (12)
	Na	399	298 (4)
	C	400	97 (32)
	C	400	299 (4)
	A	401	97 (32)
	A	401	300 (4)
	A + C	402	97 (32)
	A + C	402	301 (4)
	B	404	102 (32)
	B	404	303 (4)
	B + C	405	102 (32)
	B + C	405	304 (4)
	A + B	406	102 (32)
	A + B	406	305 (4)
	A + B + C	407	102 (32)
	A + B + C	407	306 (4)
UDP-Glc-NAc	Na	-606	-79 (60)
	Na	-606	-385 (32)
	D	-608	-79 (60)
	D	-608	-385 (32)
	A	-609	-79 (60)
	A	-609	-388 (32)
	B	-611	-79 (60)
	B	-611	-390 (32)
	A + D	-611	-79 (60)
	A + D	-611	-388 (32)
	C	-612	-79 (60)

C	-612	-385 (32)
B + D	-613	-79 (60)
B + D	-613	-390 (32)
A + B	-614	-79 (60)
A + B	-614	-393 (32)
C + D	-614	-79 (60)
C + D	-614	-385 (32)
A + C	-615	-79 (60)
A + C	-615	-388 (32)
A + B + D	-616	-79 (60)
A + B + D	-616	-393 (32)
A + C + D	-616	-79 (60)
A + C + D	-616	-388 (32)
B + C	-617	-79 (60)
B + C	-617	-390 (32)
B + C + D	-619	-79 (60)
B + C + D	-619	-390 (32)
A + B + C	-620	-79 (60)
A + B + C	-620	-393 (32)
A + B + C + D	-632	-79 (60)
A + B + C + D	-632	-393 (32)

Supplementary table 4. Labeling experiment culturing conditions. Labeled C13 glucose and glutamine composition.

13C-Glucose Complete medium

Nacalai-Tesque 09848	500 mL
13C-Glucose 1 g/l	
Sodium Pyruvate 0.11 g/l = 1 mM	5 mL from 100 mM stock
FBS	50 mL

13C-Glutamine Complete medium

Sigma M5650	500 mL
13C-Glutamine 0.292 g/l = 2 mM (stock 5 mg)	
Sodium Pyruvate 0.11 g/l = 1 mM	5 mL from 100 mM stock
FBS	50 mL

Chapter 3

Chapter 3 will be published as a preprint in biorxiv and will be submitted to Cell Host & Microbe or BMJ Gut. This work has been supported by the European Union's Horizon 2020 under the grant agreement No 675610, 824110-EASI-Genomics and the Spanish National grant BFU2017-87958-P.

9. Biological sex and microbiota drive epigenetic marks and transcription in the mice liver

9.1. Introduction

Sexual dimorphism is associated with physiological and molecular characteristics of an organism that allows us to differentiate each sex. Human sexual dimorphism exists in healthy homeostasis at different locations, times and molecular levels such as RNA expression (*613*); DNA methylation (*614*, *615*); metabolism (*616*); or protein levels (*617*), often in a tissue specific-manner (*618*). In addition, diseases also have a marked sexual dimorphism, and we can find examples of differences in sex incidence in cancer (*619*); type 2 diabetes and cardiovascular risk (*620*); obesity (*621*); inflammatory bowel disease (*622*); alzheimer disease (*622*, *623*); or autoimmune diseases (*624*) among others, giving us a glimpse of how biological sex can influence the incidence and prevalence of a disease. The challenge arises when we use animal models in the context of translational research, where we attempt to model a human disease or study a molecular aspect, in a closely related mammalian species such as the house mouse (*Mus musculus*). Usually, male animals are chosen by default in most experiments, which poses a limitation to translate the findings of the investigation to both sexes in the general population. Despite the implementation of policies and recommendations from science funding agencies such as the European Union (EU) or the National Institute of Health (NIH) trying to address this bias (*625*), (626), (627), scientists still include females less often than males in human, animal and cell research (*628–631*). Females are excluded from experimental designs based on the unfunded belief of introducing higher variability than males, neglecting half of the population (*632*, *633*). A recent pandemic meta-analysis of COVID-19 trials show that only 17.8% of published studies reported sex-disaggregated results (*634*). These alarming figures are not solely observed in human clinical trials, but also in basic, translational and pre-clinical research areas.

Microbiome has been no exception to the sex-biased research. Imbalances in the mammalian gut microbiota are associated with host health (*509*, *635*, *636*); development and aging (*637–639*); nutrition (*640*, *641*); or behavior (*35*), among other host physiological features. However, most striking observations made in microbiome research can only be attributed to one sex: males (*483*, *541*), (*511*, *642*), (*84*), (*493*), (*643*), (*644*, *645*). Females have been somewhat forgotten leading to a considerable gap in our knowledge on the crosstalk between sexual dimorphism and microbiota in health and disease. Still, a few seminal publications have shown that the commensal microbial community and sex hormone levels are reciprocally regulated, which can determine genetic predisposition to disease in a gender-specific manner (*187*), (646), (*647*), (*648*), (*183*), (*649*). For instance, only some microbial

lineages can provide the necessary signals to regulate the gender bias in T1D development (*646*). A recent study has also shown that the microbiome is required for sex-specific diurnal rhythms of gene expression and metabolism (*170*).

Despite recent works linking biological sex, host genetics and microbiota, very little is known about the molecular determinants and unforeseen relationships that could explain sexual dimorphism. Here we study germ-free and conventional mice with equal representation of both sexes in each condition, and take a multi-omics approach to further characterize and understand the host-microbiota communication by using RNA-seq, whole genome bisulfite sequencing (WGBS), proteomics of histones, targeted and untargeted metabolomics, and metallomics targeting the mouse liver (N=40), complemented with 16S metagenomics amplicon sequencing of the intestinal content from conventional mice (n=20). We found that the presence of microbiota induces DNA hypomethylation, however this hypomethylation is more pronounced in males, leading to a greater global transcriptional activation compared to the other groups.

9.2. Results

9.2.1. DNA methylation, gene transcription levels and metabolism in the liver mouse are sex and microbiota-dependent

Multi-omics studies promise the improved characterization of biological processes across molecular layers. Here, we explored our germ-free and conventional mice by using multi-omics factor analysis (MOFA), a Bayesian framework algorithm created to integrate multi-omics experiments in an unsupervised manner and discover the principal sources of biological variation (*468*) (see Methods section). We constructed the MOFA model using properly transformed values of differentially expressed genes (DEGs, n=9942), differentially methylated regions (top DMRs, n=10000), metabolites (n=4032 for untargeted, and n=32 for targeted metabolomics), metals (n=14) and histone PTMs (n=29), measured in the liver of each mouse (n=40). The first three latent factors explained most of the biological variance (> 75%) in our data (Figure 1.a). Latent factor 1 (LF1) explains >50% of the biological variance in our model, separating males from females (Figure 1.b). Yet, conventional males (CVM) show higher variance in LF1. Latent factor 2 (LF2) explained >22% of the variance and separated the germ-free from the conventional animals. Latent factor 3 (LF3) explains >4% of the variance in our model, showing that GFF are masculinized, or it can also be seen as feminized GFM, a phenomenon already described previously (*170*, *650*, *651*).

The cumulative variance decomposition by factor can be seen in Figure 1.c. Major differences between sexes (LF1) are driven by epigenetic and genomic features, that is, DNA differentially methylated regions (DMRs) and gene expression profiles (RNAseq), whereas the main differences between germ-free and conventional mice (LF2) are driven by metabolic signatures. Notably, histone PTMs and metals have a minor contribution to the latent factors, possibly due to more stable levels in the liver. We observe a clear separation of the four conditions when values for factor 1 and 2 are plotted, which indicates that methylation marks in DNA are sex and microbiota-dependent (Figure 1.d). Note how males with microbiota (CVM), are more separated from the rest in Figure 1.b and d, indicating different molecular signatures in males depending on the microbiota status.

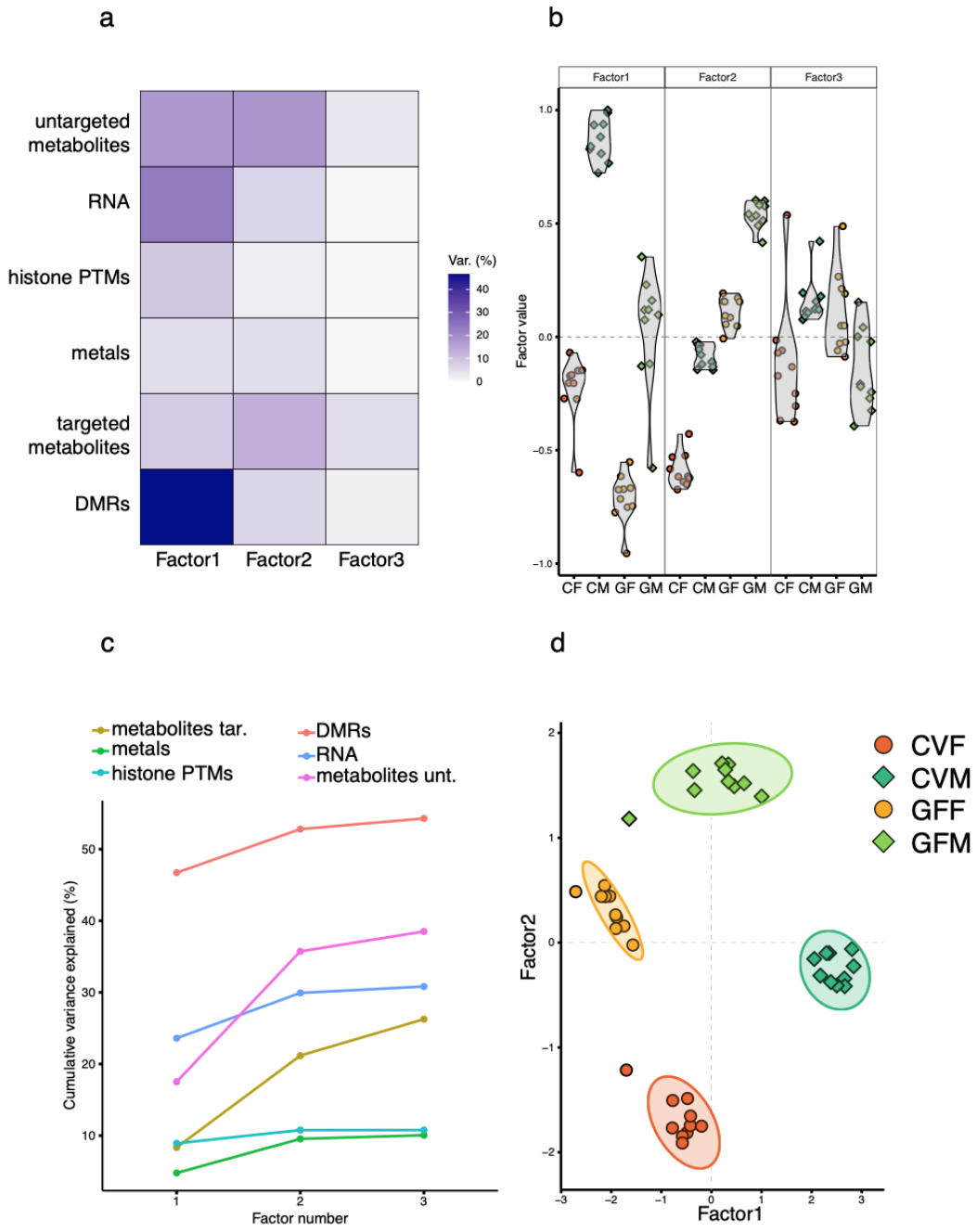


Figure 1. Multi Omics Factor Analysis (MOFA), and underlying variance explanation. MOFA main plots depicting the main source of variance in the three main latent factors. Heatmap showing the percentage of variance explained by each molecular view (i.e. omic dataset) in each latent factor (a). Cumulative variance explained in each view (b). Violin plots showing the separation of the experimental groups in the three latent factors (c). Scatterplots of factor 1 and 2 values showing the separation of the experimental groups (d).

9.2.2. RNA functional analysis suggest a sex and microbiota-dependent steroid and lipid metabolism

First, we analyzed the whole RNA dataset (>20K genes), from which the DEGs used to construct the MOFA model were extracted, as a more direct read-out of the effect of having or not microbiota, or being a male or a female in the gene expression profile and its functionality. We used weighted correlation network analysis with the WGCNA *R* package (v1.72-1) aimed to create a weighted gene coexpression network. We correlated each module of co-expression to the pertinance to each experimental group (Figure 2a). Conventional males (CM or CVM) had the higher number of significant modules correlated, and germ-free males (GM or GFM) had the lower number of significant modules correlated. When performing the overrepresentation enrichment for each module, we observed a clear influence of sex and microbiota on steroid and fatty acid metabolism. This analysis indicates that steroid metabolic process is depleted in CVM and enriched in GFF, or that the fatty acid metabolism is enriched in CVM and depleted in both female groups (CVF and GFF).

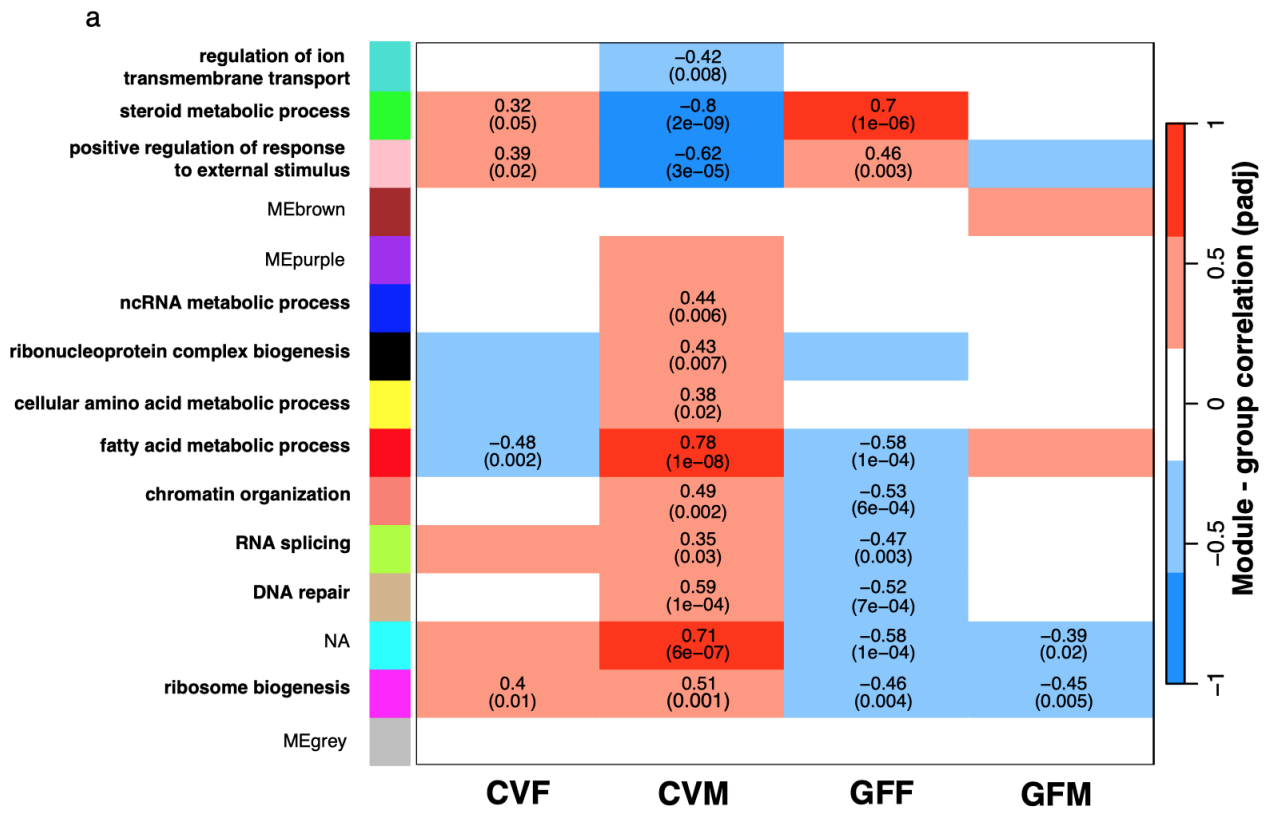
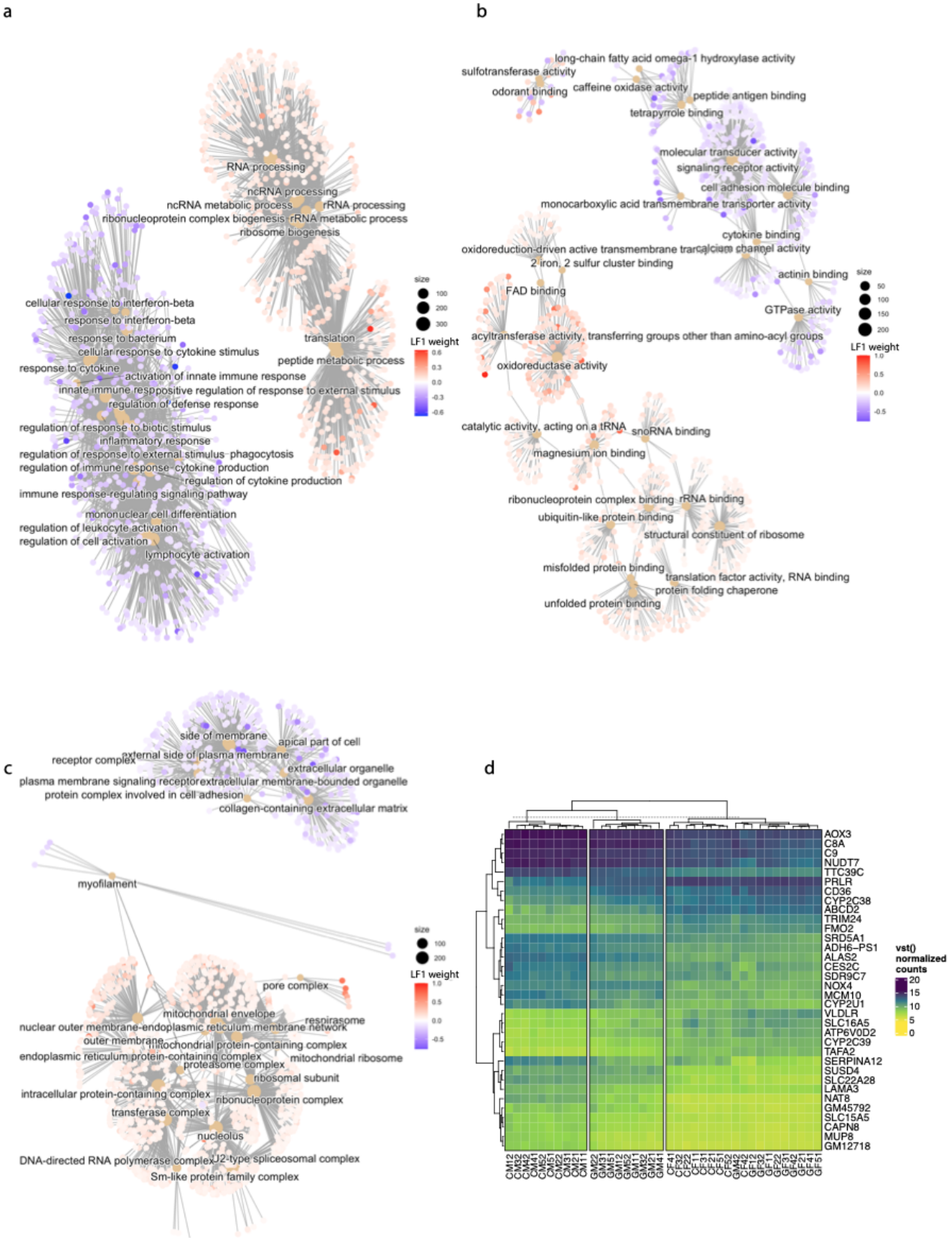


Figure 2. Weighted correlation network analysis. Correlation heatmap of all genes clustering together eigenvectors (colors or GO terms), correlated to the experimental group in the WGCNA analysis. Top one enriched with the most overrepresented term when enriched (top numbers correspond to the WGCNA correlation value, and lower values to the p adjusted (BH) < 0.05) (a).

Next, we explored further the RNA view, to look into the underlying functionality encoded in the RNA features that separate CVM from the rest in the LF1 of the MOFA model. Differences between males and females at the transcriptional level are known, and have been the object of numerous publications (*613*, *652–656*). We performed a gene set enrichment analysis (GSEA) with the *gseGO()* function from *clusterProfiler* R package, taking the scaled weights for all features in the RNA view of the model (positive and negative), analogously to the fold-change of a typical GSEA analysis to rank the genes. Positive LF1 weights were enriched in biological process such as ribonucleoprotein complex biogenesis, RNA processing or non-coding RNA processing (Figure 3a); enriched in molecular functions such as structural constituent of ribosome, iron-sulfur cluster binding or oxidoreductase activity (Figure 3b); and enriched in cellular components such as mitochondrial membrane, DNA-directed RNA polymerase complex or mitochondrial respiratory chain complex (Figure 3c), all terms being associated to CVM and partially associated to GFM. In contrast, negative LF1 weights were enriched in biological process such as inflammatory response, response to bacterium or innate immune response (Figure 3a); enriched in molecular functions such as signaling receptor activity, cell adhesion molecule binding or GTPase activity (Figure 3b); and enriched in cellular components such as plasma membrane receptor complex, extracellular organelle or apical part of the cell (Figure 3c), negative weights being associated to CVF and GFF. The positive enriched functions suggest a transcriptional activation of the CVM, to regulate reactive oxygen species, protein folding and acetylation, endoplasmic vesicle traffic, steroid hormone metabolism and the regulation of general mitochondrial activity implicated in some of those biological functions.


a
RNA processing
ncRNA processing
ncRNA metabolic process
rRNA processing
ribonucleoprotein complex biogenesis
rRNA metabolic process
ribosome biogenesis
translation
peptide metabolic process
cellular response to interferon-beta
response to interferon-beta
response to bacterium
cellular response to cytokine stimulus
response to cytokine
activation of innate immune response
innate immune response
positive regulation of response to external stimulus
regulation of defense response
regulation of response to biotic stimulus
inflammatory response
regulation of response to external stimulus
phagocytosis
regulation of immune response
cytokine production
regulation of cytokine production
immune response-regulating signaling pathway
mononuclear cell differentiation
regulation of leukocyte activation
regulation of cell activation
lymphocyte activation
size
LF1 weight
b
long-chain fatty acid omega-1 hydroxylase activity
sulfotransferase activity
odorant binding
caffeine oxidase activity
peptide antigen binding
tetrapyrrole binding
molecular transducer activity
signaling receptor activity
cell adhesion molecule binding
monocarboxylic acid transmembrane transporter activity
cytokine binding
oxidoreduction-driven active transmembrane transporter activity
calcium channel activity
2 iron, 2 sulfur cluster binding
actinin binding
FAD binding
GTPase activity
acyltransferase activity, transferring groups other than amino-acyl groups
oxidoreductase activity
catalytic activity, acting on a tRNA
snoRNA binding
magnesium ion binding
ribonucleoprotein complex binding
rRNA binding
ubiquitin-like protein binding
structural constituent of ribosome
misfolded protein binding
translation factor activity, RNA binding
protein folding chaperone
unfolded protein binding
size
LF1 weight
c
side of membrane
apical part of cell
receptor complex
external side of plasma membrane
extracellular organelle
plasma membrane signaling receptor complex
extracellular membrane-bounded organelle
protein complex involved in cell adhesion
collagen-containing extracellular matrix
myofilament
pore complex
mitochondrial envelope
respirasome
nuclear outer membrane-endoplasmic reticulum membrane network
outer membrane
mitochondrial protein-containing complex
endoplasmic reticulum protein-containing complex
mitochondrial ribosome
proteasome complex
ribosomal subunit
intracellular protein-containing complex
ribonucleoprotein complex
transferase complex
nucleolus
DNA-directed RNA polymerase complex
U2-type spliceosomal complex
Sm-like protein family complex
size
LF1 weight
d
AOX3
C8A
C9
NUDT7
TTC39C
PRLR
CD36
CYP2C38
ABCD2
TRIM24
FMO2
SRD5A1
ADH6–PS1
ALAS2
CES2C
SDR9C7
NOX4
MCM10
CYP2U1
VLDLR
SLC16A5
ATP6V0D2
CYP2C39
TAFA2
SERPINA12
SUSD4
SLC22A28
LAMA3
NAT8
GM45792
SLC15A5
CAPN8
MUP8
GM12718
vst() normalized counts

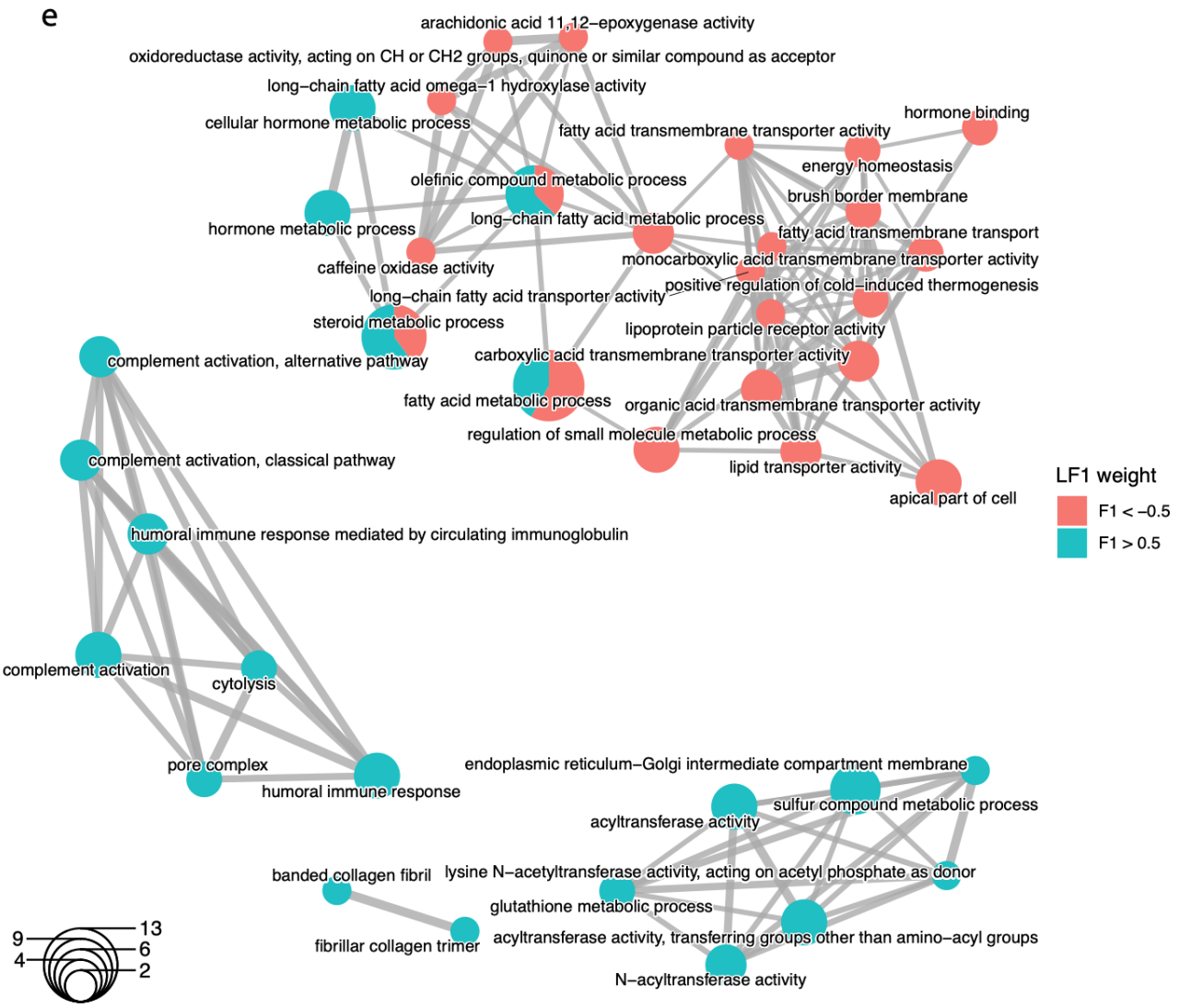


Figure 3. Mitochondrial biology, lipid metabolism and immune response influenced by sex and microbiota. Results for the GSEA and ORA performed in the RNA features of LF1. Biological processes (a), molecular functions (b) and cellular components (c) enriched in the gene set enrichment analysis with weights in the LF1 as ranking score. Cutoff for significance in each enrichment was set to p.adj <0.05 (BH). Heatmap of the top 30 genes in the LF1 (d). Emapplot with the ORA enrichment results for the RNA features that had an absolute weight higher than 0.5 in the MOFA Factor 1 (LF1), to highlight and capture the main functions separating the CVM from the rest, red color corresponding to negative associated to CVF, GFF and to a less extend GFM, and blue corresponding to positive weights associated to CVM (e).

To complement the GSEA analysis we performed an overrepresentation test (ORA) with features weighting over 0.5 or under -0.5 in the LF1 (76 genes, from which 4 mitochondrial). Considering positive weights, associated to CVM, and negative weight to the other groups. The normalized counts of the top 30 genes are represented in Figure 3d, showing marked differences between the 4 groups, but especially in CVM. Top positive features (47 genes, from which 2 mitochondrial) were enriched in biological process such as steroid metabolism, sulfur compound metabolism or cellular hormone

metabolism; enriched in molecular functions such as N-acyltransferase activity, sulfur compound binding or glutathione metabolism; involving cellular components such as the pore complex, endoplasmic reticulum-Golgi membrane or banded collagen fibril. Top negative features (29 genes, from which 2 mitochondrial) were enriched in biological process such as fatty acid metabolic process, small molecule metabolism regulation or response to xenobiotic stimulus; enriched in molecular functions such as lipid transport activity, hormone binding or cytokine binding; involving cellular components such as apical part of the cell, receptor complex or brush border membrane (Figure 3e).

The apparent discordance between the GSEA and ORA analysis is because the GSEA distincts small, but coordinated changes in a list of ranked genes, while the ORA aims to find major overrepresented biological functions based on the presence or absence of the gene in the term, missing small but coordinated changes in gene expression (*657*). GSEA was performed using all weights associated with all genes used in the MOFA model, while the ORA was performed using only top weighted genes (i.e. abs(weight) > 0.5) to highlight the functions enriched in CVM in respect to the rest. We can extract from the ORA analysis that the main biological functions that separate CVM from the rest are related to fatty acids and steroid metabolism. Putting together both analysis shows biological functions enriched in mitochondrial located or interacting proteins, trafficking vesicles between organelles in the cell, regulation of transcription and immune response to bacteria.

Mitochondria associated membranes (MAM) vesicles are trafficking lipids (i.e. cholesterol) from endoplasmic reticulum (ER) to mitochondria (*658*). For example, in the synthesis of steroid hormones, the first and rate-limiting step to convert cholesterol to pregnenolone takes place in the mitochondria (*659*). Steroid metabolism cytochrome P450 (CYP) type I proteins in general, are heme-containing enzymes located in the inner mitochondrial membrane (*660*). And the iron-sulfur metabolism is related to the synthesis of the heme group, and is an important reactive oxygen species producer and mediate electron transfer in the mitochondrial respiratory chain (*661*). In our RNA enrichment results, top terms like sulfur metabolism, steroid metabolism, cytolysis, mitochondrial respiratory chain or oxidoreductase activity are linked to mitochondrial activity, suggesting a sex and microbiota-dependent regulation of mitochondrial biology.

9.2.3. **Additive DNA methylation changes by male sex and presence of microbiota in mouse liver**

Next, we investigated the DNA methylation view where we detected over 21M CpGs representing 97% of the total CpGs present in the mouse genome. The detected CpGs sites are located in five main genomic regions: promoters (18% of the CpG sites), exons (4%), introns (36%), UTRs (2%) and distant regulatory elements (41%). Among the 21M detected CpGs, over 52K CpGs (representing 0.25% of the total) fell into >11.8K differentially methylated regions (DMRs) with an average of 6 CpGs per DMR, defined as genomic regions with different DNA methylation status across any of the four biological conditions (Figure 4a).

We found the levels of DNAm in intronic regions of CVM group significantly lower when compared to the methylation levels in nearby exons or untranslated regions (UTRs). However, performing the same comparisons in CVF, GFM or GFF groups, had no significant differences, suggesting a specific sex and microbiota-dependent hypomethylation of intronic DMRs (Figure 4g).

Over 60% of the annotated DMRs were close to the transcription starting site (TSS), that is, within 20k bases up or down the TSS. Interestingly, while CVF, GFM or GFF had a pronounced hypomethylation pattern flanking (i.e. less than 10kb up or down) the TSS, CVM in contrast, were overall less methylated and had a constant methylation pattern regardless of the distance (Figure 4h), which possibly indicate a major accessibility of regulatory regions and their genes by transcription factors, promoting a higher transcription in males with microbiota.

We then inspected the proportions of total DMRs detected in each comparison, to detect if the additivity of sex and microbiota was global or due to a specific group. Noticeably, the percentage of DMRs was over 70% of total DMRs, when comparing methylation levels between GFM or CVF to the methylation levels of CVM (Figure 4i). In contrast, when comparing the methylation level differences between GFF and CVF, or GFF and GFM, represented around 10% or 15% from the total DMRs detected, respectively. Suggesting CVM to be the group with higher changes in DNA methylation when compared to the other groups and due to a sex-microbiota additivity effect.

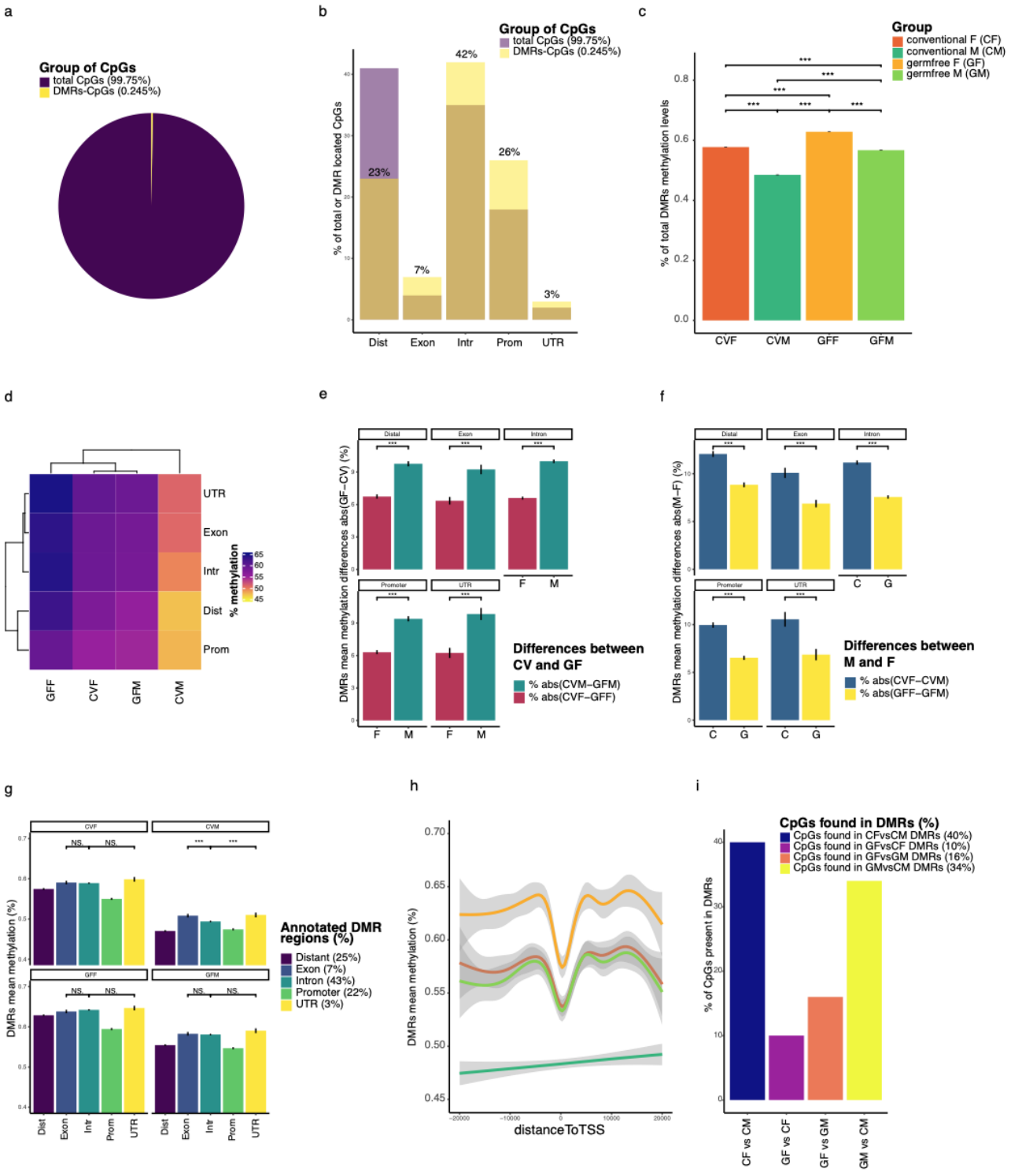


Figure 4. Male hypomethylation in the different DMRs regions. Pie chart with the total CpGs detected and in yellow the percentage of CpGs falling into DMRs (a). Proportion of total CpGs by annotated region and proportion of CpGs falling into a DMR (b). Global mean of B values showing the DMRs methylation levels per experimental group (c). Heatmap of mean B values for DMR methylation status by group and annotated genetic location (d). Barplot showing the mean DMRs methylation levels changes between males (germ-free and conventional) or females (germ-free and conventional)(e). Barplot showing the mean DMRs methylation levels changes between conventional (males and females) or germ-free (males and females)(f). B values of DNA

methylation in the 5 annotated regions and showing only the comparison that only differs in CVM (g). Scatter plot showing the fitted smooth line to highlight the constant levels of hypomethylation near the transcription starting site (TSS) of CVM with respect to the other groups (h). Percentage of DMRs from the total calculated DMRs corresponding to each comparison (i). (Wilcox p.adj < 0.05: *; < 0.01: ** ; < 0.001:***)

Male sex and microbiota have an additive effect on DNA methylation, dominated by the influence of sex. Figure 5a-f shows the distribution of DMRs in GFM, CVF and CVM relative to GFF with the measured (y) or predicted (x) of CVM/GFF values. Remarkably, when considering a linear model with an independent term, DNA hypomethylation changes in CVM can be predicted with very high accuracy from the sum of GFM and CVF relative to GFF (x axis, corrected by a multiplicative factor), suggesting some complex regulation at the epigenome scale by the crosstalk of male sex and microbiota activity.

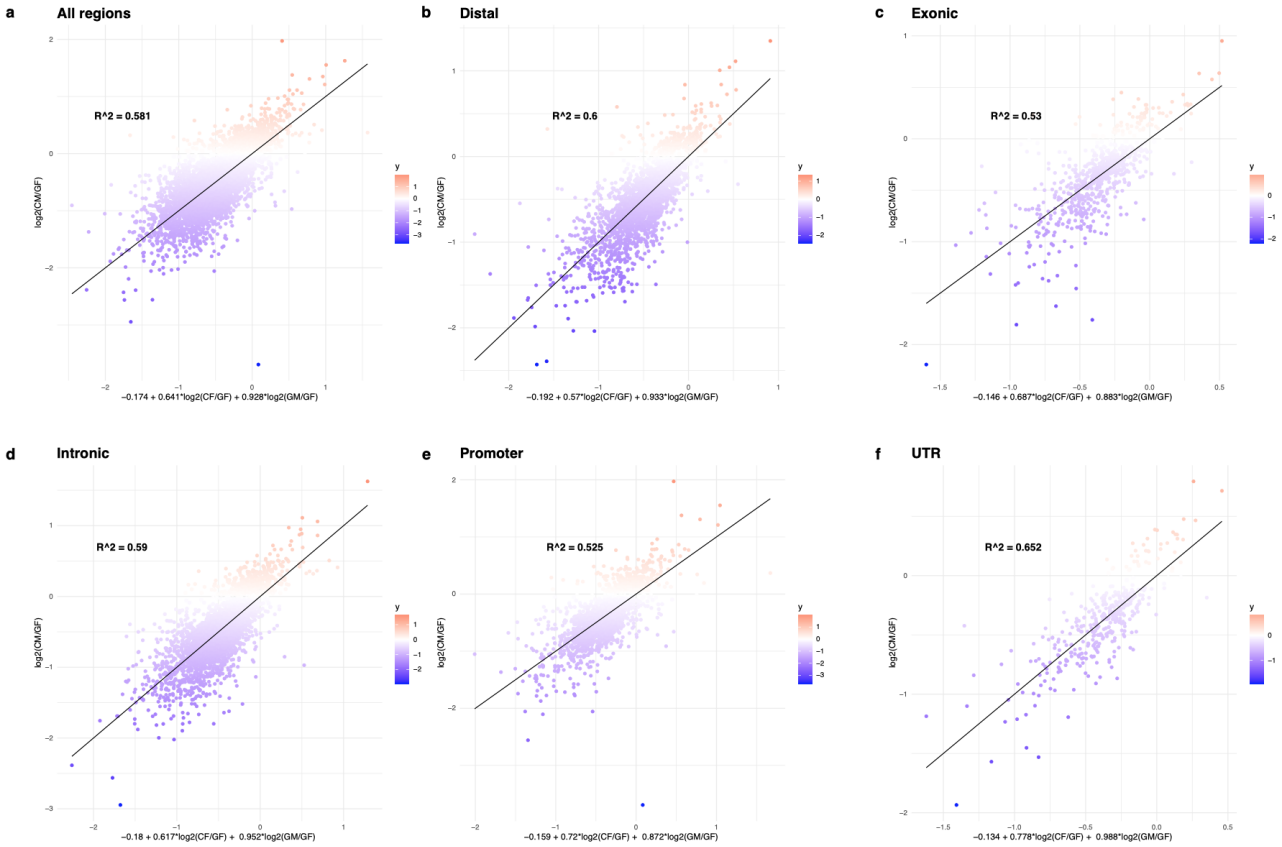


Figure 5. DNA methylation male sex and microbiota additivity linear model. Linear model fitted with the fold changes between methylation levels of CVM and GFF, and the sum of fold changes between methylation levels in CVF and GFM, with respect to levels in the GFF. Represented in the scatter plot the measured fold change (CVM/GFF) in y axis, and the predicted value of fold change CVM/GFF with its corresponding equation in x axis, after fitting the linear model for all regions (a). Measured (x) versus predicted (y) CVM/GFF values in distal (b), exonic (c), intronic (d), promoter (e) or untranslated regions (UTR)(f).

9.2.4. **Additive DNA methylation changes are mainly associated with the synthesis of steroid hormones**

Next, we explored the functional implications of the additive effect of sex and microbiota activity on DNA hypomethylation and hypermethylation. Hypomethylated regions in CVM (Figure 6a), are associated with lipid metabolism, including fatty acids, steroid/sterol and cholesterol metabolic processes.

Hypermethylated regions in CVM relative to GFF (Figure 6b), are associated with ureteric bud and mesonephric tubule development, which are key processes for the development of the urinary and reproductive organs. In this sense, monooxygenase activity, mainly steroid hydroxylase activity through cytochromes P450 (CYP450s), which promote the biosynthesis of steroid hormones, is also associated with hypermethylated regions in CVM. In the heatmap of Figure 6c we show how sample groups are clustering driven by a gradient in the methylation levels from hypermethylated GFF→CVF→GFM→CVM to hypomethylated, visualizing again the additivity of sex and microbiota, guiding the levels of DNA methylation.

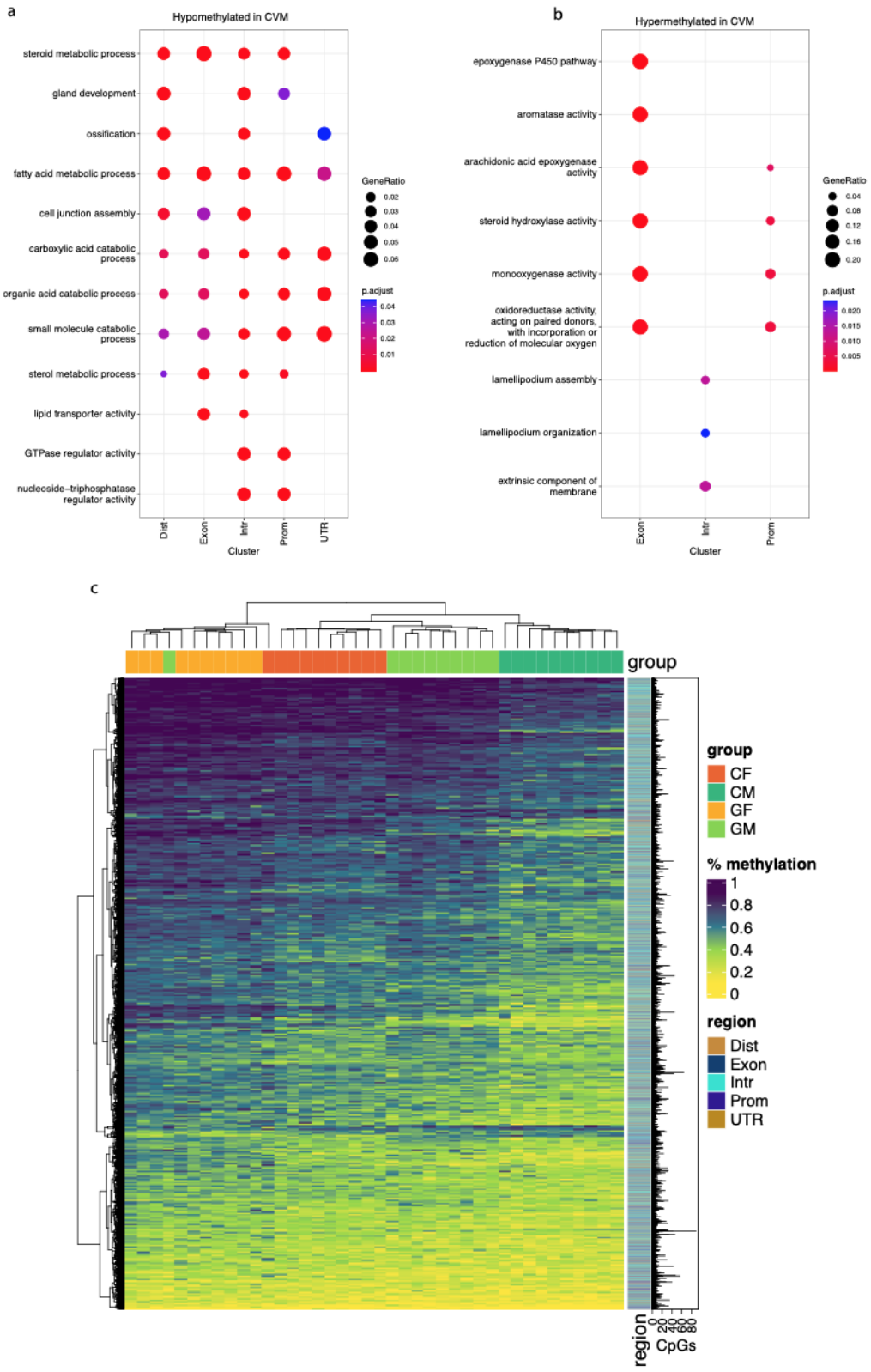


Figure 6. Sex and microbiota additivity functional enrichment and heatmap. Enrichment of hypomethylated DMRs in males with microbiota (a), or hypermethylated DMRs (b). Heatmap of methylation ß values of hyper and hypomethylated DMRs from the linear model of Figure 5a(c).

9.2.5. **Steroid hormones and lipid metabolism functions enriched in strong DMR-RNAm correlations**

DNA methylation status and the levels of gene expression have been described to have a negative correlation, meaning more methylation in the gene body or its regulatory regions, correlated to lower levels of gene expression (*252*, *662–664*). However, the opposite has also been described, where higher methylation levels in some specific regions have been correlated to more gene expression (*665*, *666*).

To investigate the impact of DNA methylation on gene expression and cellular functions in germ-free and conventional males and females, we used RNAseq data to calculate the Pearson correlation between all differentially expressed genes (9942 DEGs) and DMRs (See methods). Out of 99M possible correlations, we focused on very strong correlations (abs(r) > 0.85), which corresponded to >44k DEG-DMR pairs involving 656 unique DEGs and 1610 unique DMRs (associated with 1203 unique genes). Figure 7a shows the number of DMRs correlated with each DEG, while Figure 7b shows the number of DEGs correlated with each DMR gene (defined as the closest gene to the DMR). Interestingly, the large majority of the 656 DEGs correlated positively (355) or negatively (434) to a wide range of DMRs (from 1 to >650), but we rarely detected DEGs with both positive and negative correlations with DMRs at the same time (Figure 7a). In contrast, DMRs tend to correlate positively and negatively to a similar number of DEGs, indicating a pleiotropic effect of each individual DMR (Figure 5b).

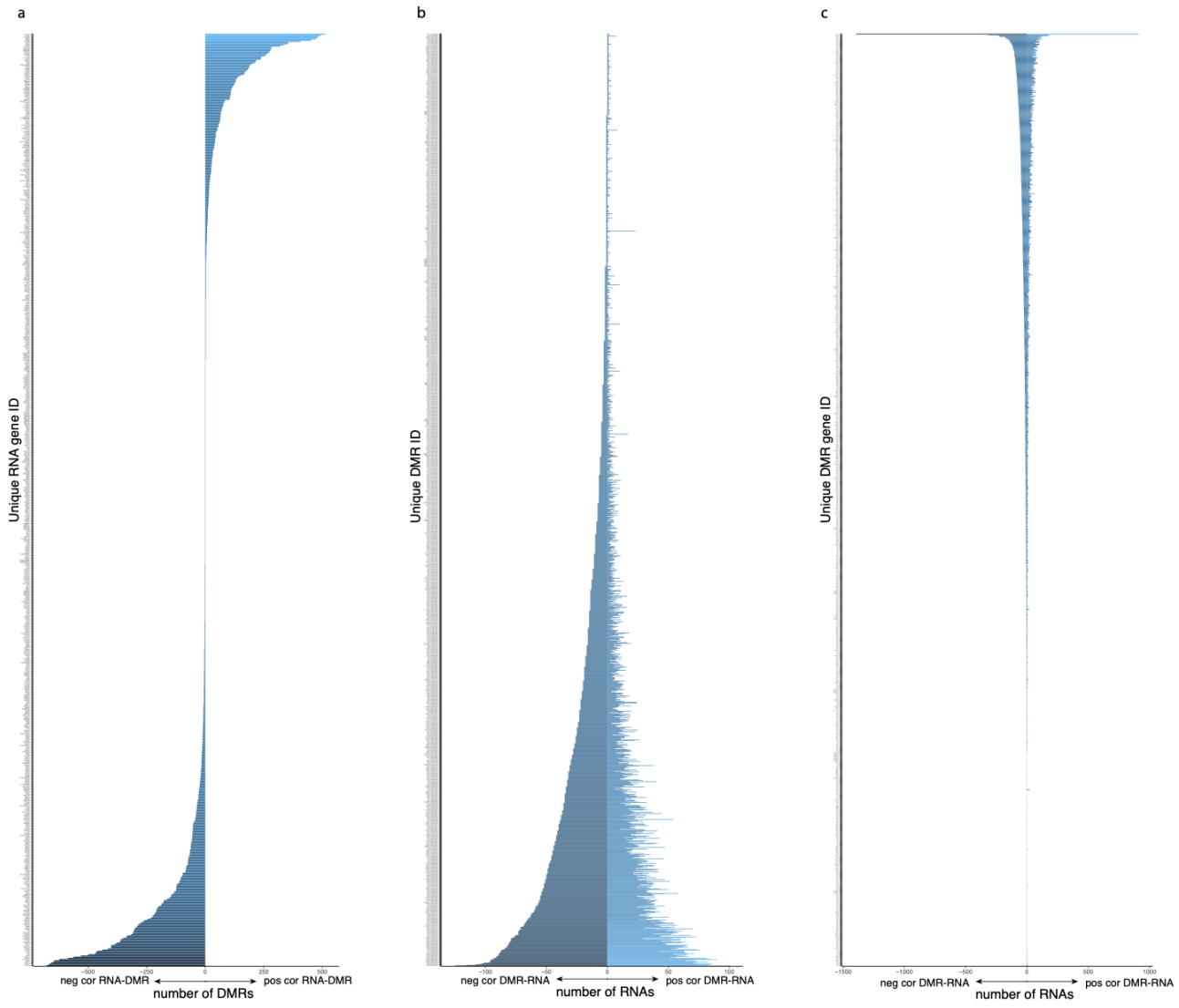
On average, each DEG has 34 strongly correlated DMRs, while each DMR gene correlates on average with 21 different DEGs. Interestingly, the cytochrome P450 superfamily of enzymes and major urinary proteins (Mups) DEGs, concentrates the largest number of correlated DMRs, while Tetratricopeptide Repeat Domain 39C (*Ttc*39c) is the protein coding gene with more associated DMRs (50) and, by far, the DMR gene with more correlated DEGs, in particular, 1396 and 913 negatively and positively correlated DEGs, respectively (Figure 7c). ATP Binding Cassette Subfamily B Member 10 gene (*Abcb*10) has a DMR that regulates/correlates with the expression of 225 RNA DEGs, being the individual DMR that correlates with more genes. Negative correlations represented 63% of the strong correlations and 37% were positively correlated. Surprisingly, both positive and negative strong correlations were overrepresented in the same or similar biological terms when performing the enrichment with the gene IDs annotated near the DMRs, or when performing the enrichment using the RNA gene IDs (Figure 7d-e). Terms related to steroid hormones, lipid metabolism, sulfur metabolism, intracellular transport, monooxygenase activity or cellular response to xenobiotic were among the most abundant in both, positive and negative correlations.

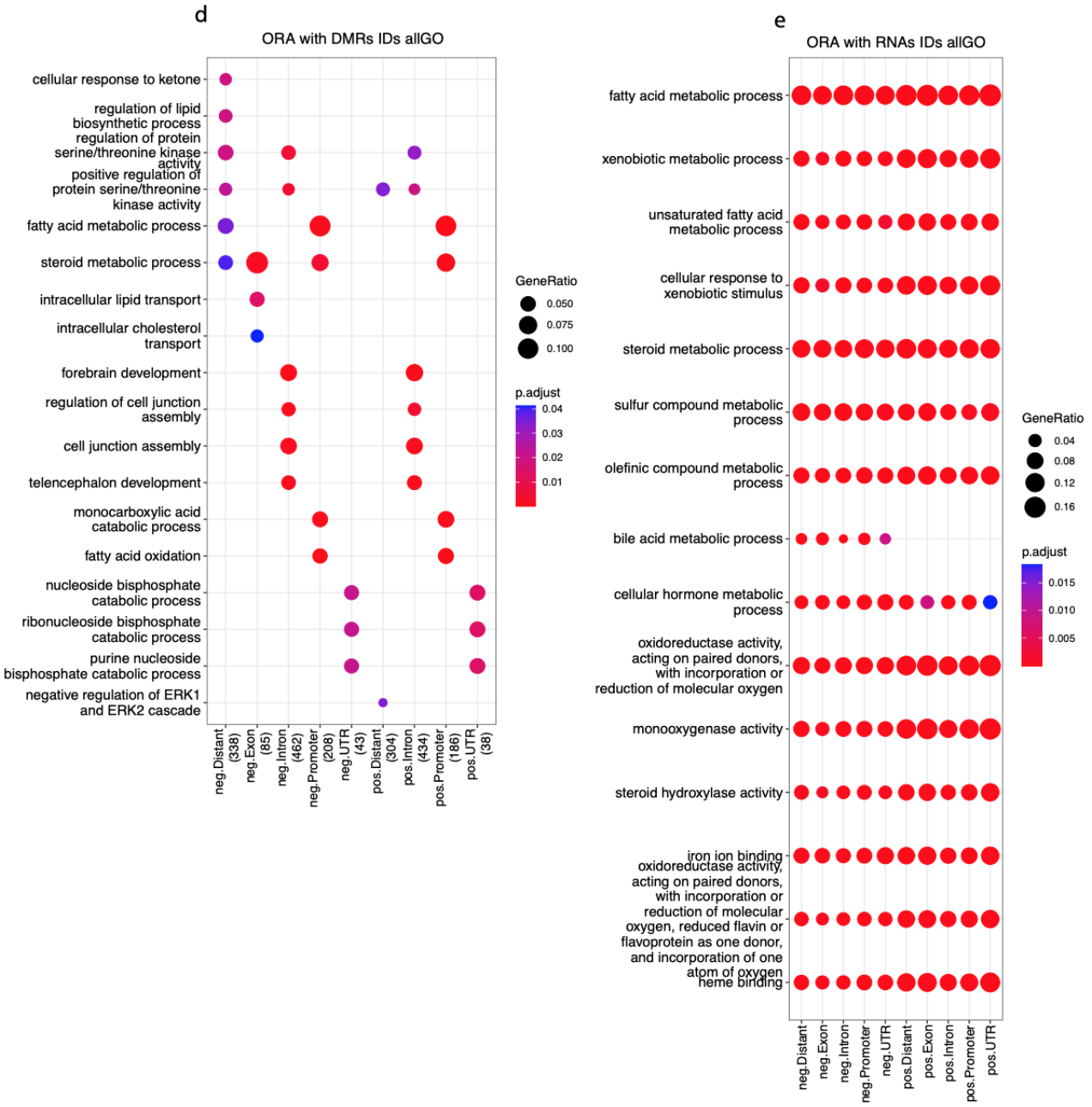
However, when separating the enrichment by the sign of the correlation and annotated region, we observed a more granulated picture. Using the gene IDs near annotated DMRs, the ORA enrichment shows an interesting profile where each biological function is controlled at the same time by positive and negative correlations. Astonishingly, each enriched function when using the associated DMR gene, was only correlated with one or few types of genomic location (Figure 7d). For example, the term fatty acid oxidation was positively and negatively correlated to the methylation levels in promoter regions. In contrast, very few terms were correlated to more than one genomic region, suggesting those terms

being actively fine tuned by the methylation levels in different genomic locations at the same time, being those terms related to lipid metabolism, steroid hormones and cell signaling. Steroid metabolism was negatively correlated with the methylation levels in distal, promoter and exon regions, and positively correlated only with methylation in promoter regions. Fatty acid metabolism was very similarly correlated, with the exception that it was not negatively correlated to exon methylation. Interestingly, methylation levels in the exonic regions had terms negatively correlated, but no terms were positively correlated to the levels of DNA methylation in exons.

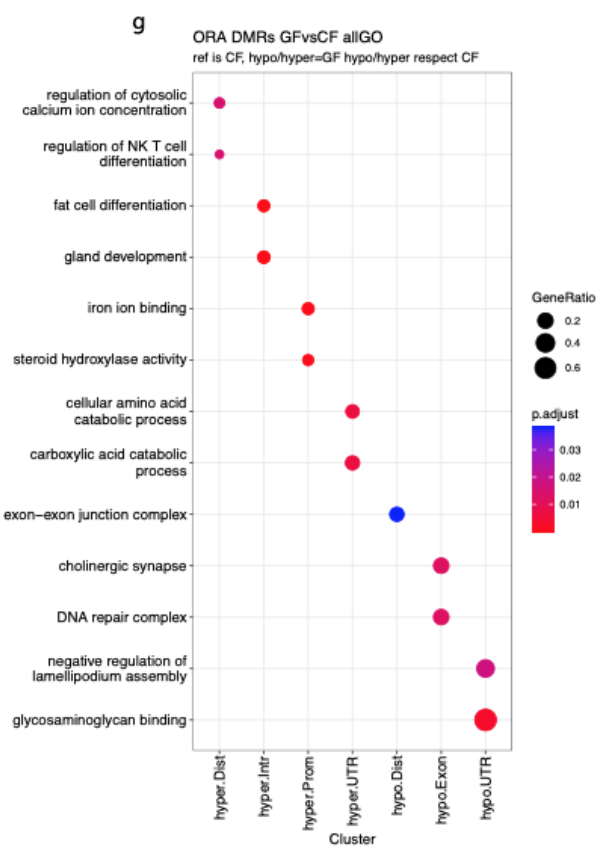
However, when performing the enrichment analysis with the gene IDs from the RNA features (Figure 7e), all top functions were positively and negatively correlated to the levels of DNA methylation in all genomic regions, except the bile acid metabolic processes term. Top terms are again related to fatty acid metabolism, steroid hormones, iron-sulfur metabolism, monooxygenase activity with the addition of bile acids metabolism. Bile acid metabolism encoded in RNA genes is the only term not positively correlated to the methylation levels of any genomic region as appreciated in Figure 7e, indicating a specific epigenetic mechanism regulating this function based on DNA methylation that would need further investigation. Bile acids are known steroid compounds that undergo secondary conjugation processes mediated by bacteria in the gut and with specific immunomodulatory functions (*34*, *95*). In Figure 7d-e we used very strong correlations between DEGs and DMRs, and performed the enrichment with the gene IDs annotated near the DMR (Figure 7d) or with the DEGs gene IDs (Figure 7e).

When performing the overrepresentation enrichment with the DMR genes from the pairwise comparisons CVF vs CVM, GFF vs CVF, GFM vs CVM or GFF vs GFM (Figure 7f-i), we can observe that in all comparisons, hypomethylated and hypermethylated regions are enriched in different functions in a methylation and DMR-region dependent manner. Indicating the control of different functions by hypo- or hypermethylation of DMRs located in different genomic regions. We can notice that for example terms related to fatty acid metabolic process, other lipid related terms or carboxylic acid catabolic process are hypomethylated in most genomic regions of CVM when compared to CVF or GFM (Figure 7f,i). However, terms related to membrane dynamics in CVM are hypermethylated and have an exclusive location of the DMRs when compared to GFM. For example the apical plasma membrane term is only controlled by DMRs located in intronic regions (Figure 7i). Interestingly, when comparing both female groups (GF vs CF), or both germ-free groups (GF vs GM), one or few functions were controlled exclusively by DMRs located in one genomic region (Figure 7g-h).


a
Unique RNA gene ID
neg cor RNA-DMR
pos cor RNA-DMR
number of DMRs
b
Unique DMR ID
neg cor DMR-RNA
pos cor DMR-RNA
number of RNAs
c
Unique DMR gene ID
neg cor DMR-RNA
pos cor DMR-RNA
number of RNAs


d
ORA with DMRs IDs allGO
cellular response to ketone
regulation of lipid biosynthetic process
regulation of protein serine/threonine kinase activity
positive regulation of protein serine/threonine kinase activity
fatty acid metabolic process
steroid metabolic process
intracellular lipid transport
intracellular cholesterol transport
forebrain development
regulation of cell junction assembly
cell junction assembly
telencephalon development
monocarboxylic acid catabolic process
fatty acid oxidation
nucleoside bisphosphate catabolic process
ribonucleoside bisphosphate catabolic process
purine nucleoside bisphosphate catabolic process
negative regulation of ERK1 and ERK2 cascade
neg.Distant (338)
neg.Exon (85)
neg.Intron (462)
neg.Promoter (208)
neg.UTR (43)
pos.Distant (304)
pos.Intron (434)
pos.Promoter (186)
pos.UTR (38)
GeneRatio
0.050
0.075
0.100
p.adjust
0.04
0.03
0.02
0.01
e
ORA with RNAs IDs allGO
fatty acid metabolic process
xenobiotic metabolic process
unsaturated fatty acid metabolic process
cellular response to xenobiotic stimulus
steroid metabolic process
sulfur compound metabolic process
olefinic compound metabolic process
bile acid metabolic process
cellular hormone metabolic process
oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen
monooxygenase activity
steroid hydroxylase activity
iron ion binding
oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, reduced flavin or flavoprotein as one donor, and incorporation of one atom of oxygen
heme binding
neg.Distant
neg.Exon
neg.Intron
neg.Promoter
neg.UTR
pos.Distant
pos.Exon
pos.Intron
pos.Promoter
pos.UTR
GeneRatio
0.04
0.08
0.12
0.16
p.adjust
0.015
0.010
0.005


f
ORA DMRs CFvsCM allGO
ref is CM, hypo/hyper=CF hypo/hyper respect CM
regulation of protein serine/threonine kinase activity
regulation of cell–cell adhesion
intracellular lipid transport
very long–chain fatty acid metabolic process
fatty acid metabolic process
carboxylic acid catabolic process
organic acid catabolic process
GTPase regulator activity
nucleoside–triphosphatase regulator activity
monooxygenase activity
epoxygenase P450 pathway
long–chain fatty acid metabolic process
oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, reduced flavin or flavoprotein as one donor, and incorporation of one atom of oxygen
arachidonic acid epoxygenase activity
steroid hydroxylase activity
hydrolase activity, acting on carbon–nitrogen (but not peptide) bonds
phosphopyruvate hydratase activity
hyper.Dist
hyper.Exon
hyper.Intr
hyper.Prom
hyper.UTR
hypo.Exon
hypo.Intr
hypo.Prom
hypo.UTR
Cluster
GeneRatio
0.05
0.10
0.15
p.adjust
0.04
0.03
0.02
0.01


g
ORA DMRs GFvsCF allGO
ref is CF, hypo/hyper=GF hypo/hyper respect CF
regulation of cytosolic calcium ion concentration
regulation of NK T cell differentiation
fat cell differentiation
gland development
iron ion binding
steroid hydroxylase activity
cellular amino acid catabolic process
carboxylic acid catabolic process
exon–exon junction complex
cholinergic synapse
DNA repair complex
negative regulation of lamellipodium assembly
glycosaminoglycan binding
hyper.Dist
hyper.Intr
hyper.Prom
hyper.UTR
hypo.Dist
hypo.Exon
hypo.UTR
Cluster
GeneRatio
0.2
0.4
0.6
p.adjust
0.03
0.02
0.01


h
ORA DMRs GFvsGM allGO
ref is GM, hypo/hyper=GF hypo/hyper respect GM
positive regulation of lipid biosynthetic process
regulation of lipid biosynthetic process
synapse organization
regulation of synapse organization
small molecule catabolic process
carboxylic acid catabolic process
inositol trisphosphate kinase activity
testosterone 16–beta–hydroxylase activity
positive regulation of mucus secretion
regulation of mucus secretion
actin filament
collagen trimer
extracellular matrix structural constituent conferring tensile strength
hyper.Dist
hyper.Intr
hyper.Prom
hypo.Exon
hypo.Intr
hypo.Prom
hypo.UTR
Cluster
GeneRatio
0.1
0.2
0.3
0.4
0.5
p.adjust
0.020
0.015
0.010
0.005


i
ORA DMRs GMvsCM allGO
ref is CM, hypo/hyper=GM hypo/hyper respect CM
steroid metabolic process
gland development
fatty acid metabolic process
regulation of lipid metabolic process
lipid homeostasis
cell junction assembly
carboxylic acid catabolic process
organic acid catabolic process
GTPase regulator activity
mesenchymal stem cell proliferation
positive regulation of epithelial cell apoptotic process
regulation of digestive system process
regulation of gastric acid secretion
apical part of cell
apical plasma membrane
negative regulation of collagen metabolic process
negative regulation of collagen biosynthetic process
integrin binding
cell adhesion molecule binding
hyper.Dist
hyper.Exon
hyper.Intr
hyper.Prom
hyper.UTR
hypo.Dist
hypo.Exon
hypo.Intr
hypo.Prom
hypo.UTR
Cluster
GeneRatio
0.05
0.10
0.15
0.20
p.adjust
0.04
0.03
0.02
0.01

Figure 7. RNA-DMR strong correlations and functional analysis. Number of DMRs strongly correlating (abs(cor) >0.85) to each unique DEG (a). Number of RNAs strongly correlating (abs(cor) >0.85) to each unique DMR (b). Number of RNAs strongly correlating (abs(cor) >0.85) to each unique gene associated/near a DMR (c). Overrepresentation analysis (ORA), using gene IDs annotated near the DMRs in strongly correlated (abs(cor) >0.85) DMR-RNA features (d). ORA using RNA gene IDs in very strong correlated DMR-RNA features (e). Overrepresentation analysis (ORA), using the DMRs gene IDs extracted from the following pairwise comparison CVF vs CVM (f, 4961 DMRs), GFF vs CVF (g, 1032 DMRs), GFF vs GFM (h, 1910 DMRs), and GFM vs CVM (i, 3905 DMRs).

9.2.6. **Low abundant *Melainabacteria* class strongly correlates to the expression of mitochondrial genes**

Next, to investigate the association of bacterial taxa abundance present in the gut with the levels of gene expression in the liver, we performed the correlation analysis between the present gut taxons in the colon and the liver expressed genes. In this correlations we can mainly find two different scenarios related to a single taxa, simple correlations were a single taxa correlates to simple or few features in one of the other data modalities (i.e. one bacterial taxa correlates to one or few expressed genes or putative metabolite/s), or multiple correlations were a single taxa correlates with several features (i.e one bacterial taxa correlates to many DMRs or expressed genes). We have selected two examples, one to illustrate how a low abundant single bacterial taxa (*Melainabacteria* class) is correlated to the levels of very few differentially expressed genes; and a second example, to show how two high abundant bacterial related taxa (one uncultured species and one uncultured genus from the *Ruminococcaceae* family, belonging to the *Clostridia* class) are correlated to the level of methylation of many DMRs and expressed genes (see section 9.2.7).

Melainabacteria are a rare/low abundant (less than 1% in our data, Figure 8a) gut bacterial class belonging to the *Cyanobacteria* phylum that inhabit the mouse gut lumen and are in higher abundances in CVM compared to CVF (Figure 8b). These differences suggest the occupation of a very specific ecological niche that is sex-dependent in the colon by this low abundant class, pointing to having an essential role for the mouse holobiont (host + microbiota). It has been recently shown that low abundance taxa exhibit higher immunogenicity than higher abundance taxons (*667*), or being the drivers of the bacterial community composition (*668*). Healthy male mice had a higher relative abundance of *Melainabacteria* compared to healthy females (Figure 8a-b), and two key questions arise from this observation, what do they do? and why healthy males and females have such different relative abundance of this taxa?

To answer the first question, *Di Rienzi* and colleagues discovered that *Melainabacteria* are a non-photosynthetic, nitrogen fixators, flagellated and obliged anaerobes *Cyanobacteria* class in the mammalian gut (*669*). They reconstructed their genome and found a large number of genes encoding lipopolysaccharide biosynthesis, typically found in Gram-negative bacteria with cell-envelope. The authors also found that are obligate anaerobes producing a variety of fermentation products such as lactate, acetate, formate or butyrate, compounds belonging to or closely related to short chain fatty acids (SCFAs). This *Cyanobacteria* occupy a very specific ecological niche being syntrophic H_2-producers and obligate anaerobes in the gut of mammals with a plant-based diet, including

non-westernized humans (*670*). The reconstructed genomes of *Melainabacteria* has also shown complete pathways for B and K vitamins production, such as riboflavin, nicotinamide or dihydrofolate (*669*), pointing to a key role in the physiology and homeostasis of these vitamins in the mouse holobiont, by this specialist and low abundant taxa.

First, in order to investigate the effect of this class in host gene expression, we calculated the Pearson correlation between this bacterial class and genes that were differentially expressed in conventional males only (224 genes), females only (76 genes), or in opposite directions in males and females (62 genes) that were also significant in the interaction of the two experimental factors, sex and microbiota status. We found strong correlation (abs(r)>0.85) between the abundances of the *Melainabacteria* class and expression levels of 35 genes. Some of these genes were the aldehyde oxidase 3 (*Aox*3), mitochondrial DNA polymerase gamma 2 (*Polg*2, Figure 8e), insulin-like growth factor 1 (*Ifg*1), N-acetyltransferase 8 (*Nat*8, Figure 8f), carboxyl esterase 1 epsilon (*Ces*1*e*, Figure 8g), oligodendrocyte transcription factor 1 (*Olig*1, Figure 8h), glutathione s-transferase pi 1 (*Gstp*1), a probable mannosyltransferase (*Dpy*19l3) and few other genes. Being key genes in steroid metabolism (*Ces*1*e*), in mitochondrial replication (*Polg*2), with a possible role in histone acetylation (*Nat*8), enabling RNA polymerase II cis-regulatory region sequence-specific DNA binding activity (*Olig*1), in reactive oxygen species production or oxidative stress protection (*Aox*3 and *Gstp*1).

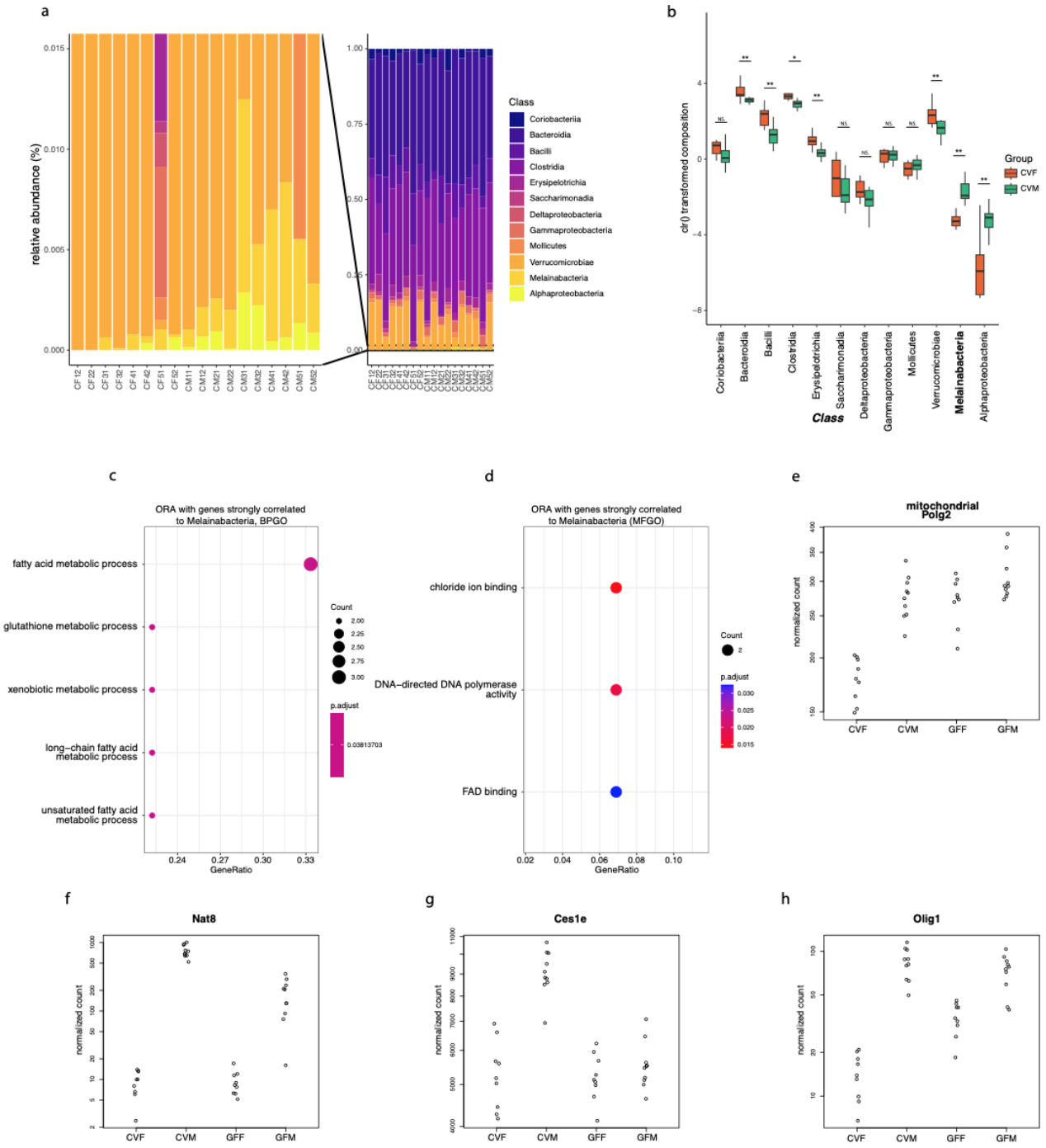


Figure 8. Relative abundances of bacterial classes and liver RNA-gut *Melainabacteria* correlations. Relative abundans barplot of all the classes present in the colon of conventional mice (a). Boxplot of the transformed relative abundances to highlight that males have higher abundances of only 2 classes (Wilcox p.adj < 0.05)(b). Dotplots showing the biological processes (BP) and molecular functions (MF) for RNA genes with a very strong correlation with *Melainabacteria* (c and d). CountPlots of normalized counts for the *Polg*2 (h), *Nat*8 (f), *Ces*1*e* (g) and *Olig*1 (h).

We then noticed in our correlation analysis between the *Melainabacteria* and the mentioned DEGs, the presence of important mitochondrial genes such as *Polg*2, *Ces*1*e*, *Tstd*1, *Acsm*2 or *Gstp*1 which can be located in the mitochondria, as well as genes related to vesicle trafficking between organelles and signaling pathway such as *Emc*9, *Rassf*3 or *Fzd*4. The ORA enrichment analysis with the 35 DEGs that strongly correlated to *Melainabacteria* revealed enrichment in functions related to fatty acids metabolism, glutathione metabolism, response to xenobiotic stimulus, DNA-directed DNA polymerase activity or FAD binding, suggesting *Melainabacteria* being key stimulators of fatty acid metabolism and mitochondrial activity in the liver (Figure 8c-d).

Second, we explored the correlation between liver untargeted metabolites and the *Melainabacteria* class. Surprisingly, the only strong positive correlation in conventional males (abs(r)>0.85) with *Melainabacteria* was a putative metabolite corresponding to a diglyceride (DG(20:3n9/0:0/18:2n6)). At the same time, the putative liver diglyceride was strongly correlated with expression of genes involved in steroid and lipid metabolism such as *Ces*1*e* and *Aox*3, or related to signal transduction such as *Gstp*1. Pointing *Melainabacteria* to have a potential role in the signaling, activation or regulation of lipid metabolism and mitochondria in conventional males. In contrast females *Melainabacteria* lower abundances were correlated to a more diverse set of putative metabolites such as few dipeptides, a phosphoethanolamine (PE) and a lysophosphatidylethanolamine (lysoPE), among other putative metabolites, possibly involved in signaling and lipid metabolism.

9.2.7. High abundant *Ruminococcaceae* family members are negatively correlated to DNA methylation levels

Next, to show how a single higher abundant bacterial taxa correlates to many variables in the RNA expression or the DNA methylation data, we used an uncultured/unidentified species and genus from the *Ruminococcaceae* family. Both species and genus are likely to be the same taxonomic entity but with small differences in their correlations with the DMRs and DEGs. From the 320 strong correlation pairs (abs(r)>0.85) between colon taxons and DEGs, 180 pairs involved one of the two *Ruminococcaceae* family members, representing over half of the correlations related to this family (Figure 9a). From the 405 strong correlation (abs(r)>0.85) pairs between colon taxons and DMRs, 217 pairs involved one of the two *Ruminococcaceae* family members, being more than half of the correlations related to these family members (Figure 9b-c).

The *Ruminococcaceae* family is a known high butyrate producer, a type of SCFAs with anti inflammatory effects (*671*), being depleted in patients with celiac diseases or ulcerative colitis (*672*). The base levels of the *Ruminococcaceae* family have been linked to antibiotic-associated diarrhea, where patients with lower abundances at the baseline were more likely to develop this pathology (*673*).

Next, we investigated the functionality encoded in the RNA-*Ruminococcaceae* correlations, being 42% negative correlations and 58% positive correlations (Figure 9a). The functional enrichment in Figure 9d shows that the abundances of the *Ruminococcaceae* family negatively correlated to the expression of genes involved in steroid metabolism, oxidoreductase activity and testosterone degradation. While positive correlations were enriched in the expression of genes involved in the complement activation, immune response, and fatty acid metabolism. Pointing to a strong link between the surge of

testosterone observed in males, the repression of gene expression involved in degrading testosterone and the *Ruminococcaceae* family members. Indicating an epigenetic repression mechanism through the hypermethylation of 4 DMR genes related to testosterone degradation (Figure 9e).

In contrast, above 96% of the correlations between the methylation levels in the DMRs with the *Ruminococcaceae* members were negative correlations, where more bacterial abundance corresponded to lower levels of methylation in the DMR (Figure 9b-c). Around 4% of the correlations were positive correlations, where more *Ruminococcaceae* abundance correlated to more methylation levels in the DMRs, or vice versa. Surprisingly, from 107 DMR genes correlating to *Ruminococcaceae* members, only the 4 genes mentioned were correlated positively, from which 3 genes were related to steroid metabolism (Cyp450 cluster) and one gene involved in SCFAs/anion transmembrane transporter. Unexpectedly, no term was enriched from over 100 DMR genes negatively correlated to *Ruminococcaceae.* Positively correlated enriched terms using DMRs genes, matched some of the terms negatively correlated to the levels of gene expression, terms related to steroid metabolism, oxidoreductase activity and degradation of testosterone. Suggesting those functions controlling the levels of testosterone by the methylation in the DMRs in the 3 genes, being more methylated in the presence of *Ruminococcaceae* and being less expressed in conventional males, where *Ruminococcaceae* was more abundant in males than in females (Figure 9f).

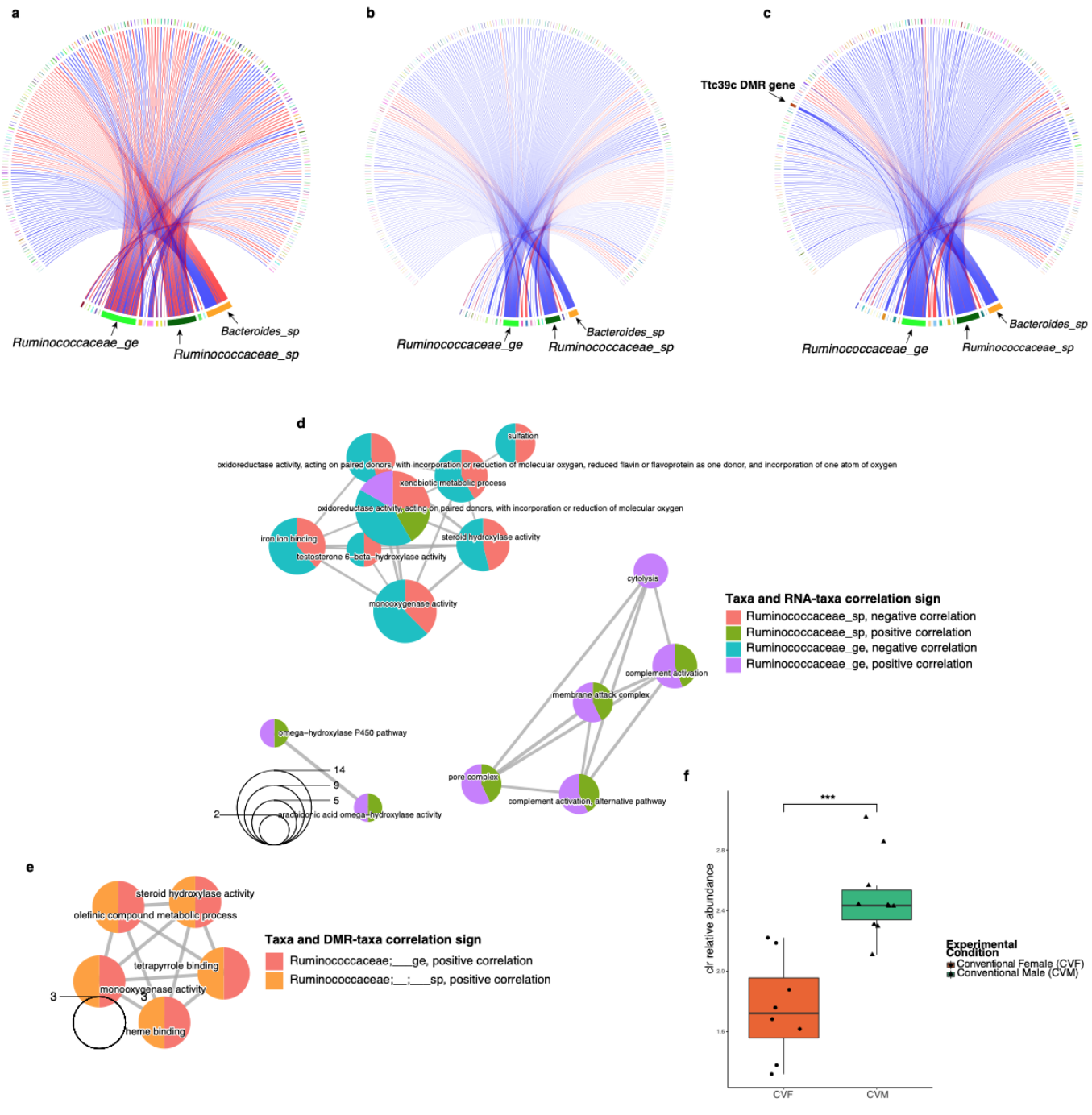


Figure 9. ***Ruminococcaceae*** **family members correlate to many DMRs and gene expression levels.** Circosplot showing the strong correlations (abs(cor)>0.85, 320 pairs) between taxons present in the colon and levels of DEGs in the liver, being red links positive correlations and blue links negative correlations (a). Circosplot showing the strong correlations (abs(cor)>0.85, 405 pairs) between taxons present in the colon and levels of methylation in individual DMRs in the liver (b). Circosplot showing the strong correlations (abs(cor)>0.85, 405 pairs) between taxons present in the colon and levels of methylation in genes near the DMRs in the liver (c). Emapplot with the ORA enrichment results using the RNA gene IDs of the strong correlations between colon *Ruminococcaceae* and levels of gene expression (d). Emapplot with the ORA enrichment results using the DMR gene IDs of the strong correlations between colon *Ruminococcaceae* and levels of methylation in the DMRs (e). Boxplot showing the abundances of the unidentified/uncultured *Ruminococcaceae* genus in conventional females and males (f)

9.3. Discussion and concluding remarks

Observing our results, it is tempting to partially attribute some of the findings to the different abundance of taxons such as *Melainabacteria* or *Ruminococcaceae*, more abundant in males than in females. But the truth origin in the differences between the four groups is undoubtedly linked to the sex and microbiota additivity effect on DNA methylation, and to a lesser extent the RNA expression. Where males with microbiota are the group with a more hypomethylated DNA, and where germ-free females are the group with a more hypermethylated DNA, having similar levels of DNA methylation in conventional females and germ-free males. These observations are coordinated with the number of differentially expressed genes in each group, being conventional males the group with more associated DEGs, and germ-free females the group with less associated DEGs in total. For the first time, we show here the additive effect of sex and microbiota in the levels of liver DNA methylation that drive the different expression profiles in the mice liver when the microbiota is present or absent, and depending on the sex.

The levels of hypo- or hypermethylation in the DNA might be driven by an equilibrium between the gut microbiota composition and its ecosystem dynamics, together with its relationship with the host and its metabolism. Our findings point to an equilibrium between low abundances specialized taxons in the gut, such as *Melainabacteria* and *Alphaproteobacteria*, and more generalist-high abundance taxons such as *Clostridia* (including the *Ruminococcaceae* family), *Bacilli* or *Bacteroidia* as it has been recently suggested (*131*), to be partially responsible for the DNA methylation levels. This ecological equilibrium is also partially responsible for the differences in gene expression observed between healthy males and healthy females mice, when the microbiota is present. The regulatory mechanisms of gene expression that we propose for further investigation, in the context of the nutrient-microbe metabolism-host epigenetic axis (*22*), is to identify microbes that correlate, stimulate, target or regulate the DNA methylation levels of genes involved in steroid and lipid metabolism. Where these microbes would have the potential use as probiotics, aiming to modulate host epigenetic marks and help with patient care or disease prevention.

The results we presented here, could partially explain why germ-free mice are resistant to a high fat diet (HFD) induced obesity and present altered cholesterol metabolism (*674*). Or partially explain why females increase the infiltration of T_{reg} in adipose tissue, conferring them resistance to obesity under a HFD (*675*). A recent prospective human study highlighted how a diet intervention and changes in the microbiota composition through the intervention, impacted the immune profile of the studied subjects (*207*). Gut microbiome has been pointed to have an essential role in the tumor treatment efficiency through the recruitment of T cells in mice and humans (*73*). Related to lipid metabolism, another study has found that lipids in the tumor microenvironment have a contradictory role modulating the immune response, activating or inactivating immunosuppressive functions of T cells in a context dependent manner (*676*). However, a possible mechanism of how the microbiota influence lipid metabolism and immune functions through epigenetic modulation of the host, has not been yet proposed.

Strong evidence exists that intercellular communication between organelles is needed for proper response to external stimuli, healthy aging and regulation of general cell homeostasis (*677*, *678*). The mitochondrion is an important cell organelle that helps with the responses to environmental stimuli,

rather than only being the powerhouse of the cell. It has essential roles in endocytosis and mitochondrial transfer (*679*), modulate the immune response by altering ROS, TCA cycle or fatty acid oxidation (*680*), or respond to dietary and microbially produced metabolites (*681*). Because mitochondrial functions are controlled by two genomes, the mitochondrial and the nuclear, it has been proposed that methylation of the nuclear DNA can affect mitochondrial functions related to aging decline (*682*). Then, it is imaginable that gut microbes could interact with their host through altering mitochondrial dynamics in host cells, by modifying the methylation levels in nuclear encoded mitochondrial genes, or genes that interact with the mitochondria.

The nuclear genome and the mitochondrial genome in eukaryotic cells have evolved together for at least the past 1.8 billion years (*683*). They communicate at many molecular levels, to control cell homeostasis, from intracellular organelles communication, to epigenetic gene control. A coordinated cell signaling and transporting of nuclear encoded genes to the targeted organelle, for in-situ organelle transcription is essential to coordinate mitochondria-ER communication in lipid metabolism (*684*). Our results point to the modulation of fatty acid and steroid metabolism when the microbiota is present, through the alteration of DNA methylation of genes controlling both metabolisms. The idea that gut microbiota can affect epigenetic marks in the host nuclear genome and related histones, has emerged in the biomedical research community, attributing to the mammalian gut microbiota the transcriptional control of some circadian sexual dimorphism genes (*170*), or the modulation of epigenetic marks in histones in different tissues (*25*).

We know that conventional male mice undergo a global DNA demethylation process driven by the steroid hormone testosterone around 8 weeks of life (*685*), and there is recent evidence that the microbiome composition influences the production of testosterone (*648*). Our results suggest an integration of both messages, the regulation of steroid hormones and the DNA demethylation process to be sex-and-microbiota dependent. Having the sex of the animal and the microbiota condition, an additive effect on the methylation levels. We observed in our data a global hypomethylation in nuclear DNA in males with microbiota that distincts this group from the rest. Indicating a sex-specific epigenetic-microbe-mediated underlying mechanism controlling the DNA methylation levels and gene transcription. In particular, we have observed an enrichment of DMRs present in intronic regions, similarly to what has been observed previously (*685*)(*686*). In our methylation data, we have observed intronic regions being particularly hypomethylated in males with microbiota (Figure 4g), when compared to methylation levels in exons or UTRs, but these differences were not observed in the other three groups. Suggesting that an intronic demethylation is an essential epigenetic event for conventional males maturity and sexual dimorphism, more pronounced in the presence of microbiota.

A recent study demonstrates how enhancers-like sequences (ELS) are more abundant in intronic regions involved in regulating tissue-specific functionalities, highlighting the importance of genomic location of these regulatory elements (*687*). Other cis-regulatory elements have also been reported to be located in intronic regions (*688–690*). Accordingly, and linking the DMR methylation level to the location of these ELS, nearly half of the DMRs regions we found in our methylation data were located in intronic regions. This goes in line with the findings by Borsari et al., where they observed that specialized tissues (i.e. liver) tend to have more ELS in the gene body than in intergenic regions, facilitating the access to transcription and regulatory factors. In this regard, our data shows that around

43% of the DMRs are located in intronic regions, which might help their action as regulatory elements being more accessible to the transcription machinery. Additionally, intragenic DNA methylation has been studied for the possible role that plays in alternative splicing, functioning as alternative promoter (*691*), and for sensing the levels of intragenic DNA methylation levels (*692*). Together with the general hypomethylation levels of all DMRs in the annotated regions of males with microbiota (Figure 4d), these findings support the observation that the CVM group is more transcriptionally active than its germ-free counterpart and both female groups.

Remarkably, we found one gene with 50 DMRs associated (*Ttc*39*c*), correlating with and having a possible pleiotropic effect on over 2000 DEGs, suggesting this DMR gene being a master regulator of gene transcription, which needs to be further investigated. We observed the *Ruminococcaceae* family correlating to many DEGs and DMRs, positively and negatively correlating to several DEGs, and mostly correlating negatively to many DMRs. Suggesting the *Ruminococcaceae* family as an important modulator of DNA methylation and gene expression related to steroid metabolism. However, we could not propose the exact molecular mechanism underlying these correlations, which warrants further investigation. Surprisingly, when matching gene IDs from the DEGs and the DMRs that strongly correlated to the *Ruminococcaceae* family, we found only 10 genes with the same gene ID. Three of these genes are known important steroid metabolism players which help degrading testosterone, suggesting the *Ruminococcaceae* family plays a key role in host and microbiota steroid metabolism.

In conclusion, we have observed for the first time that the DNA methylation levels in the mice liver are sex and microbiota-dependent, having females higher methylation levels compared to males, and being the microbiota responsible for the hypomethylation observed in conventional animals compared to their germ-free counterparts. The different DNA methylation levels observed between sexes (in both microbial conditions, germ free and conventional) are driven by the testosterone surge in males (*685*), being responsible for most of the methylation variability between groups. Furthermore, these differences are enhanced by the presence of microbiota and the different microbiota composition between males and females, as a second factor driving the different levels of DNA methylation observed. Particularly, the different abundances between males and females in the *Ruminococcaceae* family, seem to play an important role in the epigenetic control of testosterone degradation.

9.4. Contributions

The project was conceived and designed by Oscar and myself. The samples were obtained through a collaboration we started with Sylvie Rabot from INRAE where she leads a germ-free mouse facility. The samples were collected by Aurélie Balvay and Claire Maudet with the help of Sylvie Rabot. I performed the lyophilization, pulverization and homogenization of the liver samples, as well as aliquoting and weighting all the sub-aliquotes used for every omics and matrix analyzed. I performed the measure of most epigenetically relevant metabolites in liver, serum and intestinal content, with the exception of SCFAs that were analyzed in the center for omics science (COS, Reus, Spain). At that time, the method presented in Chapter 2 was not ready and we outsourced these measures. With the help of Ignasi Forne and Magdalini Serafidu under the supervision of Axel Imhof, I performed the histone digest extraction and preparation for the mass spectrometry analysis. Ignasi Forne generated the histone PTMs data and I analyzed the data. Tamara García Barrera analyzed the metals in the liver. Toni Gabaldon, Pedro González-Torres and Maria Belen Carbonetto generated and analyzed the 16S data from colon and cecum samples of the conventional group. I extracted, purified and measured the concentration of total RNA in all liver samples. The total stranded RNA sequencing and data generation was performed in the Centro Nacional de Analisis Genomicos (CNAG, Barcelona, Spain). Under the supervision of Marta Melé and the help of Luisa Santus, I preprocessed, mapped, aligned and obtained a matrix of counts for the RNA seq data in an internship in the Barcelona Supercomputing Center (BSC-CNS, Barcelona). Teresa Rubio, Salva Casaní and Ana Conesa helped me with the initial multivariate statistical analysis, data exploration and data integration of several omics. I wrote a proposal and obtained free-sequencing services for the whole genome bisulfite sequencing for DNA methylation analysis. Sequencing was performed by the EASI-Genomics partner DKFZ sequencing facility in Heidelberg. DNA methylation data was processed and analyzed by Christian Heyer under the supervision of Matthias Schlesner. Jordi Rofes has helped me with part of the correlational analysis, and part of the additivity plots, as well as processing the untargeted metabolomics data set. Jordi Capellades has helped me with the statistical analysis, and teached how to efficiently program in the R language. I gathered, corrected batch effects if necessary and transformed all data types, using several R packages to explore individual data types, perform univariate and multivariate statistics, and integrated all data types to generate most of the figures in the manuscript, presented in Chapter 3. I wrote the draft manuscript and Oscar and I edited until the current version.

9.5. Materials and methods

Germ-free and conventional mice model

Mouse housing

Nine-week-old female and male germ-free (GF) and conventional (CV) C57BL/6J mice were used for the experiment. The GF mice were procured by the breeding unit Anaxem, the GF facility of the Micalis Institute (INRAE, Jouy-en-Josas, France; Anaxem license number: B78-33-6). The CV mice (males and females) were purchased from Charles River Laboratories (L'Arbresle, France) and kept in the Anaxem facilities. The GF mice were housed in sterile isolators (Getinge, Les Ulis, France), and in individual cages. Fresh defecations were used to ensure sterile conditions weekly by microscopic examination and screening cultures. CV mice were housed in the same type of isolators but non-sterile, to ensure the same environment and stress between experimental groups. In each isolator, mice were kept in enriched home cages containing paper towels, wooden sticks and sterile bedding made of wood shavings, and free access to autoclaved tap water and γ-irradiated (45 kGy) standard chow diet (R03; Scientific Animal Food and Engineering, Augy, France). The experimental procedures were performed in accordance with European guidelines for the care and use of laboratory animals and approved by the ethics committee of the INRAE Research Centre at Jouy-en-Josas (approval reference: 17-14). Ten mice were used for each sex and microbiota status, with a total n=40 mice. All mice were housed individually for measurement of individual food consumption. Body weight and food were weighed once a week from the age of 4 to 9 weeks. And animal room temperature was maintained at 20–24°C with a strict 12-h light/dark cycle (lights open at 7:30 am).

Mouse necropsy and biopsies

The mice were anesthetized with isoflurane and killed by decapitation. Serum was obtained by collecting around 500 uL of blood after decapitation, followed by centrifugation at 2000 rpm for 10 min at room temperature. The supernatant was transferred in a cryo tube and frozen. Liver was extracted and put in a cryo tube and immediately frozen. The gastrointestinal tract was extracted from the body and carefully opened to collect the content in a cryo tube and frozen. Cecal and colon content were collected from the lumen of the colon and cecum and obtained from the fecal material. All samples were immediately frozen in liquid nitrogen after being collected and stored at -80°C until the corresponding analysis was performed. The serum, colon or cecum content were not further processed before the different analysis performed on those samples. However, the frozen livers were lyophilized, pulverized and homogenized to have an homogenous representation of the liver. The pulverized liver was kept at -80ºC, before proceeding to extract and analyze the nucleic acids, metabolites, metals and proteines.

RNA-seq analysis on liver samples

Total RNA extraction

An average of 12 mg of lyophilized and pulverized mouse liver was transferred into an RNase-free tube. Total RNA was extracted using the Direct-zol™ RNA Miniprep Plus (Zymo Research, Irvine, CA 92614) following the manufacturer's protocol, and performing the optional DNase I treatment. Briefly, the digestion, pre-wash and wash solutions were prepared. 600 uL of TRI Reagent were added to each liver sample, and homogenized with a vortex until the solution had a clear appearance, followed by an incubation on ice for 30 min. The sample was centrifuged at 12.800 rpm for 1 min and 500 uL of supernatant were transferred into a new RNase-free tube and discard any remaining debris. 500 uL of 100% ethanol were added and mixed thoroughly with a vortex followed by 30 min incubation on ice. 500 uL of solution were transferred into a Zymo-Spin IIICG Column with its corresponding collection tube and centrifuged to allow the RNA to bind to the column, and discarded the flow-through. The last step was repeated until the whole solution had passed through the column for each sample. 400 uL of Wash Buffer were added to the column and centrifuged, followed by adding 80 uL of digestion solution with DNase I and incubation of 15 min at room temperature. Two rounds of washing the column with 400 uL of Pre-Wash buffer and centrifugation were performed, followed by one round of washing with 700 uL of Wash buffer and centrifugation for 2 min to ensure complete removal of the wash buffer. Finally, the column was transferred into a new RNase-free tube and the RNA was eluted by adding 100 uL of DNase/RNase-free water and 1 min centrifugation.
RNA quantification was initially performed with NanoDrop 2000 (Thermo Fisher Scientific, Waltham, MA) to assess how much material was extracted. Absolute quantification of extracted RNA was measured using the Qubit instrument and the Qubit® RNA HS or BR Assay (Thermo Fisher Scientific, Waltham, MA). Quality control of the RNA integrity was performed by using the Agilent Bioanalyzer and RNA 6000 Nano Bioanalyzer 2100 Assay (Agilent, Santa Clara, CA). RNA integrity number (RIN) was calculated for all samples..

rRNA depletion, total Stranded RNA library preparation and sequencing

Ribosomal RNA (rRNA) is very abundant, up to 80-90% of the total RNA, and it was removed using the Illumina Ribo-Zero Plus rRNA Depletion Kit (Illumina, San Diego, CA). Briefly, rRNA is degraded by enzymatic reactions. Sequencing libraries were constructed after the rRNA depletion using the Illumina® TruSeq Stranded Total RNA Sample Preparation kit, following the manufacturer's recommendations. Strand specific cDNA, adapter ligation and PCR amplification was performed. The final library was validated on an Agilent 2100 Bioanalyzer with the DNA 7500 assay (Agilent, Santa Clara, CA). The libraries were sequenced on a HiSeq4000 (Illumina, Inc) in a fraction of a HiSeq 4000 PE Cluster kit sequencing flow cell lane, following the manufacturer's protocol for dual indexing. Image analysis, base calling and quality scoring of the run were processed using the manufacturer's software Real Time Analysis (RTA 2.7.7). Followed by the generation of FASTQ files. Above 60 million of 75 bp long PE reads for every sample were obtained.

RNA-seq data pre-processing

FASTQC (0.11.5) and *MultiQC* (v1.8) softwares were used to perform an initial quality control of the sequencing data and spot possible bias due to GC content, duplicate reads, adapter content, overrepresented sequences and other sequencing/library preparation bias. GRCm38 mouse (*Mus musculus*) Genome Reference Consortium Mouse Build 38 was used as assembly to map the RNA-seq reads. Mouse genome was indexed using *STAR* (2.7.0f) indexing function default parameters *--runMode* genomeGenerate and setting *--sjdbOverhang* to 100. Reads were mapped using the indexed GRCm38 mouse genome construct built in the previous step and *STAR* alignment default options, setting *--outSAMtype* to BAM *SortedByCoordinate*. QualiMap (2.2.1) was used to assess the quality of the mapping and spot possible issues, as well as having an overview of the percentage mapped reads and biotypes. The following parameters were specified, *-pe* (for PE reads), *-p* strand-specific-forward (for strand specific library construct protocol) and *--java-mem-size*=20G (to increase maximum Java heap memory size). Finally, the HTseq (0.11.2) software was used for transcript count using the function *htseq-count* and specifying the options *-f bam* (the format of the input file), *-r* pos (to specify that the alignment has been sorted by alignment position) and *-s yes* (to specify that the first read is aligned to the forward strand).

Differential gene expression analysis

After transcript quantification with *HTseq*, raw transcript-level read estimates (counts) were read into *R* (4.1.2) and were combined into a data matrix. In place of removing very low total counts < 10 per sample, as recommended in many RNA-seq analysis protocols, a custom and more subtil filter was applied. The filter consisted in keeping only transcripts that appeared in 80% or more samples of at least one of the four experimental groups. Followed by filtering transcripts that had a TPM > 0.1 value. The filtered transcript counts were transformed using the variance stabilizing transformation *vst()* function from the *DESeq2* (1.34.0) R package for data exploration (*435*). Top transcripts with the highest variance across all samples were used for principal component analysis (PCA) using *plotPCA()* function, to have an idea of the distribution of the sample groups. The differential expression (DE) analysis using the *DESeq2* package was performed on raw counts. The independent filtering option of *DESeq2* was enabled (default), filtering out genes with very low counts and thus unlikely to show significant evidence. *DESeq2* supplies methods to test for differential expression using the negative binomial distribution in a generalized linear model. Data-driven Bayes prior distributions are used and incorporated to calculate logarithmic fold changes and estimates of gene dispersion. The p values were calculated using the *Independent Hypothesis Weighting (IHW)* method (*693*) and adjusted for multiple testing with the BH procedure. Using the p values and the mean of normalized counts as a covariate, which are approximately independent under the null hypothesis, the package calculates weighted p values.

The DE testing design consisted of two experimental factors: having or not having microbiota (group); and being a female or male (sex). The testing design also included the interaction of the two factors similar to a two ways ANOVA formula, as well as controlling for batch as follow:

design *formula* = ~*batch* + *group* + *sex* + *group*:*sex*

To test if a gene was differentially expressed in any of the conditions, the Likelihood Ratio Test (*LRT*) was used, in contrast the Wald test was used to perform the different pairwise comparisons between experimental groups and subgroups. The IHW method to extract the calculated p values in the *results()* function was used followed by a correction for multiple testing with the BH method, in order to optimize power (*693*). After running the *DESeq()* and *results()* functions, a matrix of differentially expressed genes with their logarithmic base 2 fold change (log2FC) and adjusted *p* values, among other metrics was obtained. Significance was set to padj < 0.1 controlling for false discovery rate by the *IHW* method. A higher *padj* value than a classical *padj* < 0.05 was set because IHW is a robust data driven hypothesis weighting algorithm capable of outperforming the Benjamini-Hochber (BH) procedure alone, for controlling the false discovery rate (FDR, control of type I error).

RNA functional analysis

All genes and all significant genes, after running the DE analysis with the *DESeq2* package, were used to run gene set enrichment analysis (GSEA) and over representation analysis (ORA) tests using the *clusterProfiler* (4.2.2) package (*694*). The ORA is used to find big changes in gene expression, and GSEA will find smaller but coordinated changes. The log2 fold change and adjusted p values calculated by *DESeq2* were used to rank the transcripts and run the GSEA and ORA analyses with the *clustreProfiler* package. All Ensembl IDs were used to retrieve Entrez IDs, external gene name, description, gene biotype and GC content information for each feature and using the *R* package *biomaRt* (2.50.3) (*695*, *696*). Transcript enrichments in Gene Ontology (GO) Biological Process (BP), Molecular Function (MF) and Cellular Compartment (CC) terms were assessed with the functions *enrichGO()*, *gseGO()* or *compareCluster()* functions. The *simplify()* function with cutoff= 0.7 was used to select one representative redundant term and clarify the enrichment results. Minimum and maximum gene set enrichment sizes were defined to be 5 or 10, and 500, respectively. For visualizing the network of enriched GO terms and genes, the functions *cnetplot()*, *dotplot()* and *emapplot()* were used to plot top enriched GO terms per comparison. The comparisons made in the functional analysis are described in supplementary table ST1. A *padj* < 0.05 was considered significant after adjustment for false discovery rate with the BH procedure in the enrichment analysis.

Weighted gene correlation network analysis (WGCNA)

The *WGCNA* (v1.72-1) *R* package was used to confirm a higher number of genes expressed in CVM compared to the rest. Using *vst()* transformed RNA counts and encoding the pertinence to the experimental group in a binary matrix (1 = belonging to that group, 0 = not belonging to that group), the network was constructed. First, raw RNA counts were transformed using the *vst()* function from the *DESeq2* package. The function *pickSoftThreshold()* was used to choose a scale-free topological criterion to select a soft threshold power for the network construction. The adjacency matrix was calculated with a soft threshold power of ß=14 by using the function *adjacency(),* specifying the parameter *type=signed* for a signed network. The topological overlap measure (TOM) was calculated with default parameters. TOM assesses how much two genes are correlated based on the correlation of the other genes, using a biweight midcorrelation (median-based correlation) with the function *bicor()* of the same package. The *cutreeDynamic* and *flashClust* functions were used to construct the network and

define the modules. To reduce the number of clusters and have a more general view of which genes are correlated within the same genetic routes, the *minModuleSize* was set to 100. Once the network was constructed and genes assigned to a cluster module, the correlation between the categorical group to each cluster module eigenvector was calculated with the *bicor()* function. Then using the *enrichGO()* from *clusterProfiler*, we performed enrichment analysis of each module and assigned the top function to each module as it appears in Figure 2a.

Histones PTMs analysis on liver samples

From the same liver biopsies described previously (see "Mouse necropsy and biopsies section"), an average of 14 mg of lyophilized and pulverized mouse liver was transferred to a 1.5 mL tube. Histones were acid extracted, digested and desalted as previously described (*273*) except for the 4 first steps of the acid extraction. The first 4 steps were performed as follows:

1. Add 150 uL of 0.2 M H_2SO_4 per tube and vortex vigorously to resuspend the liver pellet.
2. Place the tube in the Bioruptor (Diagenode, Liege, Belgium) in the tube holder and set up parameters to high (30s) and off (30s), repeating this cycle 15 times in total.
3. Place the tubes in a rotator at 10 rpm and incubate at 4ºC ON.
4. Centrifuge for 30 min at 4ºC and 14000 rpm to remove insoluble tissue debris, and transfer the supernatant to a new 1.5 mL tube.
5. The TCA precipitation was followed as described in step 5b. of the Völker-Albert et al., 2018 manuscript.

The following steps described in sections "In-gel acylation and digestion of histones", "formic acid extraction of peptides", "LC-MS analysis of trypsin-digested histones", "targeted measurement of specific histone modifications" and "processing of spectra to quantify histone modifications" of the same manuscript were followed to acquire and processing the data. No isotopically labeled standards were used in this experiment as described in the section "use of isotopically labeled standards to quantify histone modifications".

LC-MS analysis and data processing

Briefly, the data was acquired using a Q Exactive HF/Orbitrap XL (Thermo Fisher Scientific, US), the quantitative analysis was performed using Skyline (4.2.0.19072), .raw files were imported into Skyline for quantitative analysis (MacLean et al., 2010). When importing the .raw files, MS1 and MS2 peaks were matched to a spectral library based on retention times and transitions. XCalibur Qual Browser (4.1.31.9) and GPMAW (5.02) were used to manually verify all imported histone peptide MS1 and MS2 peaks. A minimum of four MS2 ions were used to validate peak picking and assign accurate peptide ID to MS1 peak. The percentage of a specific peptide species was calculated using the peptide "family" method. This method consists of summing all (MS1) areas of all specific peptides belonging to the same peptide "family", and making each peptide relative to the total area for that "family." A peptide "family" is defined as a group of peptides with the same residues within a digested mix of proteins.

The *mixOmics* R package with *plotIndiv* function was used to plot the PCA and visualise principal components, detect outliers, and batch effect. The Hotelling ellipse at 99% level confidence was used to remove outlier samples falling out of the ellipse. Fold changes were then calculated using the PTM relative abundance in subgroup comparisons described in supplementary table ST1. Fold changes were tested for significant changes between groups using the *t.test* function from *stats* base R package. A Wilcoxon Rank Sum test was used to assess whether the distributions of relative abundances of PTMs between the subgroups were different from one another. A significant level of $p < 0.05$ was considered significant, and the FDR was controlled with the Benjamini-Hochberg Procedure for multiple testing.

Metabolomics

Standards and metabolomic reagents.

Methanol (MS grade), acetonitrile (ACN, MS grade), pyridine, methoxyamine hydrochloride (MA), formic acid (MS grade), N-methyl-N-(trimethylsilyl)-trifluoroacetamide (TMS) were purchased from Thermo Fisher Scientific (Waltham). All pure standards in table 1, 2,3,4,5,6-Pentafluorobenzyl bromide (PFBBr), hexane and acetone were purchased from Sigma-Aldrich (St. Louis). Acetyl-ADP-ribose was purchased from Santa Cruz Biotechnology (Dallas). Medronic acid was purchased from Agilent Technologies (Santa Clara) and ammonium acetate from Thermo Fisher Scientific (Waltham).

Metabolite extraction methods on mice samples.

Frozen lyophilized and pulverized mice liver, serum, lumen colon content and lumen cecal content were used to generate the targeted MS/MS data and measure the abundances of epigenetically relevant metabolites. Because of the nature of short chain fatty acids (SCFAs), two different extractions, sample preparations, and data acquisition were used. One for SCFA and the other for the rest of epigenetically relevant metabolites. SCFAs are challenging to analyze, and need special extraction protocols depending on the type of matrix analyzed. The rest of epigenetically relevant metabolites were extracted with the same protocol, with little variation between matrices.

Metabolite extraction and preparation

An average of 7.5 mg of pulverized liver, 25 uL of serum, 28 mg of cecal lumen content and 27.5 mg of colon lumen content were used for metabolite extraction with the same protocol with little modifications for each matrice. In general, a volume of 400 µL (or 300 µL for the serum) of acetonitrile methanol water solution (CAN:MeOH:H_2O in a 4:4:2 volume proportion) was added to the corresponding amount of sample and vortexed for 1 min. Then, the sample went into a round of three cycles: first incubated in liquid nitrogen for 30 s; sonicated for 20-30 s in a water bath; and finally vortexed for 30 s, repeating these steps up to a total of three times. Next, the sample was incubated at -20ºC for 60 min and then centrifuged 10 min at 22000 g (4ºC). Serum was only incubated at -20ºC for 60 min and centrifuged, with no freeze-thaw cycling, to allow any remaining protein or blood clot to precipitate. Finallys 100 µL of supernatant was transferred into HPLC vial insert.

SCFAs extraction and preparation

Colon and cecal content samples

An average of ~50 mg were placed into a 1.5 mL LoBind Eppendorf tube and mixed with 10 µL of internal standard mixture (BA-LAB, PA-LAB and AA-LAB) and 990 µL of methanol:water (50:50) mixture. Samples were vortexed for 5 min and centrifuged for 5 minutes at 15000 rpm and 4ºC. A volume of 80 µL of the supernatant was mixed with 10 µL BHA 0.1M and 10 µL EDC 0.25M. Then, samples were vortexed and incubated at room temperature for 1 hour in darkness. After the incubation, fecal extract was diluted by 20 folds in 50 % aqueous MeOH. 200 µL of diluted sample was extracted by 600 µL of diethyl ether though 10 minutes of vigorous shaking. Then, samples were centrifuged for 5 minutes at 15000 rpm and 4ºC. After the centrifugation, 40 µL of the upper organic layer was transferred and evaporated to the dryness by SPE-dryer at 40 ºC. Samples were reconstituted with 200 µl of methanol:water (1:1) and transferred to glass vials for analysis.

Liver samples:

An average of ~25 mg were placed to a 1.5 ml LoBind Eppendorf tube and mixed with 60 µL of internal standard mixture (BA-LAB, PA-LAB and AA-LAB) in methanol, 333 µL of H2 0:MeOH (1:1), 666 µL CHCl 3 and 200 µL of methanol and a steel bead. Samples were agitated with a bullet blender for 1 minute at speed 7. Samples were vortexed for 5 min and centrifuged for 5 minutes at 15000 rpm and 4ºC. A volume of 40 µl of the supernatant was mixed with 40 µl of water, 10 µl BHA 0.1M and 10 µl EDC 0.25M. Then, samples were vortexed and incubated at 25ºC for 1 hour. A volume of 600 µl of diethyl ether was added. Samples were vortexed for 10 min and centrifuged for 5 minutes at 15000 rpm and 4ºC. A volume of 400 µl of the upper layer was transferred to a new tube and evaporated under a stream of N2 . Samples were reconstituted with 200 µl of methanol:water (1:1) and transferred to glass vials for analysis.

Serum samples:

Samples (25 µl) were aliquoted to a 1.5 ml LoBind Eppendorf tube and mixed with 60 µl of internal standard mixture in methanol. Samples were vortexed for 1 min and centrifuged for 5 minutes at 15000 rpm and 4ºC. A volume of 40 µl of the supernatant was mixed with 40 µl of water, 10 µl BHA 0.1M and 10 µl EDC 0.25M. Then, samples were vortexed and incubated at 25ºC for 1 hour. A volume of 600 µl of diethyl ether was added. Samples were vortexed for 1 min and centrifuged for 5 minutes at 15000 rpm and 4ºC. A volume of 400 µl of the upper layer was transferred to a new tube and evaporated under a stream of N2 . Samples were reconstituted with 200 µl of methanol:water (1:1) and transferred to glass vials for analysis.

Targeted metabolomics analysis by LC-MS/MS

Extracted metabolites were separated with an Agilent InfinityLab Poroshell 120 HILIC-Z (HILIC-Z), 2.1 x 100 mm, 2.7 µm (PEEK lined) (Agilent Technologies), using 50 mM ammonium acetate with 5 µM of medronic acid solution as mobile phase A and 100% acetonitrile (ACN) as phase B. The mobile phase flux was set to 0.4 mL/min with a linear gradient elution that started at 98% B (time 0-2 min), followed by an isocratic gradient from 98% and finishing at 40 % B (time 2-9 min), then back to 98% B (time 9-9.5 min) and a holding time 3 min and a half at 98% B (time 9.5-13 min) to allow system stabilization. Ions were generated using both polarities at the same time, positive and negative

electrospray ionization (+ESI & -ESI). Spectral data was measured with a 6490 QqQ mass spectrometer (Agilent Technologies). The column temperature was set at 25ºC and the injection volume was 5µL for liver, serum and cecal or colon. The mass spectrometer parameters were: drying and sheath gas temperatures 270°C and 400°C, respectively; source and sheath gas flows 15 and 11 L/min, respectively; nebulizer flow was set to 35 psi; positive and negative capillary voltage were both set at 3000V; nozzle voltages were 1000V and -1500V, in positive and negative respectively; and iFunnel in positive HRF and LRF 130 and 100V, respectively; and iFunnel in negative HRF and LRF 110 and 60V, respectively. The ions and transitions that have been monitored can be found in supplementary table 1 in the parent m/z, 1st and 2nd m/z transitions columns, together with each collision energy (CE) used. All pure standards (for method optimization) and metabolite peak area were manually quantified with the Qualitative Analysis of MassHunter Workstation (Agilent Technologies).

Peak area from quality controls (QC) and samples were used to detect and correct instrumentation bias during the run, using the *sbcms* R package and the *QCRSC* function. The *mixOmics* R package with *plotIndiv* function was used to plot the PCA and visualize principal components, detect outliers, and batch effect. The *sva* R package was used to remove a sampling batch effect with its *ComBat* function. The Hotelling ellipse at 99% level confidence criteria was used to remove outlier samples. A non-parametric Wilcoxon Rank Sum test was applied in each subgroup comparisons defined in supplementary table ST1, and $p < 0.05$ was defined as having significant differences in the two groups distributions.

SCFAs analyzed by LC-MS/MS

The SCFAs extractions were analyzed using an LC-MS/MS setup and a chemical derivatization step of the SCFAs during the extraction/preparation of the sample was performed. Chromatography separation was performed using a Kinetex 2.6 µm polar C18 (100 Å, 100x2.1 mm) (Phenomenex, Torrance, CA, USA) with precolumn. Mobile phase A was 0.1% formic acid in water with 10 mM of ammonium formate, and mobile phase B was :0.1 % formic acid in methanol:isopropanol (9:1 v/v). The column temperature was set at 45ºC and the injection volume was 1 µL. The mobile phase flux was set to 0.3 mL/min with an elution gradient that started with an isocratic gradient from 32% of B at 0 min, to 60% of B at 4.6 min, followed by an increase up to 65% of B at time 5.5, followed by another increase up to 98% of B at 7 min, and staying at 98% of B for 2 min, and finally returning to 32% of B at 10 min, and staying at 32% of phase B for the last minute to stabilize the column. Peak integration was performed automatically by the Qualitative Analysis of MassHunter Workstation (Agilent Technologies), and absolute concentration was calculated using the labeled internal standards.
The absolute concentrations were used to explore the data using the *mixOmics R* package and *plotIndiv* function to draw a *PCA* and visualize clustering of samples, detect any batch effect or outlier. Outliers were removed using the same Hotellins criteria of 99% confidence level. The *sva* R package was used to correct a sampling batch effect using the *ComBat* function. A Wilcoxon Rank Sum test was performed to assess significance at $p < 0.05$, for all the comparisons described in supplementary table ST1 with a BH procedure to control for FDR.

Untargeted metabolomics liver analysis by LC-qToF

Metabolites were extracted from the liver by mixing 400 µl of acetonitrile, methanol and water in a 4:4:2 proportion to the lyophilized and pulverized matrix. After mixing the sample with a vortex, the mixture was frozen under liquid nitrogen for 10-20 s, then sonicated for 20-30 s and vortexed 20 s to homogenize the mixture. These steps were repeated for a total of three complete cycles. The untargeted LC-MS data was generated using a UHPLC system (1200 series, Agilent Technologies) coupled to a G6550A ESI-qToF mass spectrometer (Agilent Technologies) operated in positive and negative electrospray ionization mode (ESI+, ESI-). Metabolites were separated by hydrophilic interactions using an InfinityLab Poroshell 120 HILIC-Z, 2.1 × 100 mm, 2.7 µm (PEEK lined) (Agilent Technologies) column in both ion modes separately. Mobile phase A for the HILIC column was water (50 mM ammonium acetate) and mobile phase B was acetonitrile. The HILIC separation of the extracts was conducted under the following gradient: 0-2 min 98% of B; 2-10 min decrease to 40% of B; 10-10.5 min raise to 98% of B; 10.5-15 min 98% of B. The parameters for the electrospray ionization source (ESI) were as follow: gas temperature, 200°C; drying gas, 14 L/min; nebulizer, 35 psig; fragmentor, 175 V; and skimmer,65 V. The *m/z* acquisition range was set between 50 and 1100, acquiring 3 spectra/s.

Untargeted metabolomics data analysis

Raw files were converted to mzML with the Proteowizard tool *MSConvert* (v3.0.21034). The *RHermes* package (v0.99.0) was applied to three quality control samples to annotate them and obtain a list of putative annotations using the following parameters: 10 ppm error; 30.000 resolution; for positive mode, [M+H]+, [M+Na]+, [M+NH4]+, [M+K]+ and [M]+ adducts and for negative mode [M-H]- and [M+Cl]-; m/z filter between 50 and 1500 Da. Two 5s scan density windows were used for SOI detection and SOI were filtered with minimum 1000 intensity and a 0.8 isotopic fidelity score. SOIs found in >2/3 of the samples were merged into a consensus SOI list of annotations.

To quantify the samples, *xcms* (v3.18.0) was used: peaks were detected using the *centWave* algorithm with the following parameters: 10 ppm mass error and expected peak width range between 5-50 seconds; peaks were refined with the merge neighboring peaks algorithm using default parameters; peaks retention times were aligned with the *OBI-Warp* method using default parameters; peaks were grouped using a minimum fraction of 0.4, bin width of 5 seconds and bin size of 0.01 Da; missing peaks were filled using default parameters. Finally, the quantification matrix was filtered with the consensus SOI list and a curated list of annotated features was obtained.

Metal extraction, preparation and LC-MS/MS analysis

The first step was the preparation of organ homogenates by cryo-homogenization using a cryogenic homogenizer SPEXSamplePrep (Freezer/ 6770) during 30 s, at a rate of ten strokes/second. Next, the mineralization of samples was carried out in a microwave oven. Briefly, 0.1000 g of each sample was weighted in MiniXpress polytetrafluoroethylene (PTFE) vessels (5 mL) and 0.5 mg of a mixture containing nitric acid and hydrogen peroxide (4:1 v/v) was added. Rhodium was used as the internal standard at 1 $\mu g L^{-1}$. Samples were mineralized into a model MARS microwave oven (CEM, Matthews, USA).

After 10 min of pre-digestion, the PTFE vessels were closed and the mineralization was carried out in the microwave oven. The program used for the mineralization of samples was from room temperature ramped to 160 °C for 15 min and held for 20 min at this temperature and using a power of 400 W. Finally, the concentration of metals in the different tissues was determined by ICP-MS model 8800 Triple Quad ICP-MS (Agilent Technologies, Tokyo, Japan). All analyses were performed using three replicates. Nickel sampling and skimmer cones were employed, with a sampling depth of 10 mm. The forward power was set at 1550 W, and the gas flow rates were fixed at 15 L min^{-1} for plasma gas and 1.08 L min^{-1} for carrier gas. Reaction gases used to eliminate interferences were helium and oxygen (high-purity grade, >99.999%) and hydrogen (purity >95%). Most of the analysed elements required 4.5 mL min-1 flow-rate of helium. For selenium, a mixture of H2 (2 mL min-1) and O2 (40%) was used in MS/MS mode. Isotopes monitored were ^{51}V, ^{53}Cr, ^{55}Mn, ^{57}Fe, ^{63}Cu, ^{64}Zn, ^{65}Cu, ^{66}Zn, ^{78}Se, ^{80}Se, ^{95}Mo, ^{98}Mo, ^{103}Rh, ^{112}Cd, ^{114}Cd and ^{208}Pb with a dwell time of 0.3 s per isotope. A tuning aqueous solution of Li, Co, Y and Tl at 1 μg L^{-1} was used to tune the ICP-MS. Skimmer and sampling cones of nickel were used with a sampling depth of 10 mm.
Absolut concentrations were used for PCA dimensionality reduction to detect clustering patterns and any bias. Here, no batch effect was detected and the Hotelling's criteria was used to remove outlier samples. To perform the pairwise comparisons defined in supplementary table ST1, a Wilcoxon Rank Sum test was applied, setting significance level at $p < 0.05$ and correcting for multiple testing with the BH procedure.

16S metagenomics

Microbial 16S DNA: library prep and sequencing

The DNA from cecum and colon content was extracted using MagMAX CORE Nucleic Acid Purification Kit 500 RXN (Thermo Fisher, CA, Austin, USA), following the manufacturer's instructions. Samples were amplified using 16S rRNA V3-V4 regions specific primers:
V3-V4-Forward
5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG-3',
V3-V4-Reverse
5'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC-3').
The PCR was performed in 10-μl final volume with 0.2-μM primer concentration. The PCR cycle included: 3 min at 95 °C (initial denaturation) followed by 25 cycles: 30 s at 95 °C 30s at 55 °C, and 30s at 72 °C, and a final elongation step of 5 min at 72 °C. PCR products were purified using AMPure XP beads (Beckman Coulter, Nyon, Switzerland) with a 0.9x ratio according to manufacturer's instructions. The above described primers contain overhangs allowing the addition of full-length Nextera barcoded adapters for multiplex sequencing in a second PCR step, resulting in sequencing ready libraries with approximately 450 bp insert sizes. In brief, 5 μl of the first PCR purified product were used as template for a second PCR with Nextera XT v2 adaptor primers in a final volume of 30 μl using the same PCR mix and thermal profile as for the first PCR but with only 8 cycles. 25 μl of the second PCR product were purified with SequalPrep normalisation kit (Invitrogen, ThermoFisher Scientific, Waltham, MA, USA), according to the manufacturer's protocol. Libraries were eluted in 20 μl final volume and pooled for sequencing. Sequencing was performed using Illumina MiSeq (2×300 bp) and v3 chemistry with a loading concentration of 10 pM.

Microbial 16S data processing

Raw data demultiplexing forward and reverse reads were processed as following: Primer trimming, quality filtering, denoising, pair-end merging and phylotype calling were performed using the *Dada*2 R package (*452*); the phylogeny assessment was performed with the *Mafft* and *Fasttree* softwares. Phylotype was used to calculate the following alpha diversity metrics: community richness (observed Operational Taxonomic Units, OTU), community evenness (Pielou's Evenness), and quantitative measure of community richness (Shannon's diversity index). Phylotype and phylogenetic data were used to calculate the following beta diversity metrics: unweighted Unifrac distance (phylogenetic qualitative measure), weighted Unifrac distance (phylogenetic quantitative measure), Jaccard distance (qualitative measure) and Bray Curtis distance (quantitative measure). The taxonomic assignment of phylotypes was performed using a Bayesian Classifier trained with Silva database version 132 (99% OTUs full length sequences) (*697*). Taxon hits with less than 90% of similarity to the database were eliminated removing any low quality reads, sequencing artifacts or host reads. A scheme of the processing and analysis pipeline is shown in supplementary figure S2.

Microbial 16S data: statistical analysis

Alpha diversity comparisons were performed using Kruskal-Wallis non-parametric test. Beta diversity distance matrices were used to calculate principal coordinates analysis (*PCoA*) and to make ordination plots using R software. The significance of groups present in community structure was tested using Permanova and ANOSIM tests. Permdisp test was used to identify location vs. dispersion effects (*698*). Differential relative abundance of taxa was tested using two methods: ANCOM (Mandal et al. 2015) and Kruskal Wallis non-parametric test. After Kruskal Wallis, Conover's test with FDR Benjamini-Hochberg correction was added for pairwise comparison. A significant threshold was set at 0.05. *BiodiversityR* (2.11-1), *PMCMR* (4.3), *RVAideMemoire* (0.9-7) and *vegan* (2.5-5) packages were used for the different statistical analysis mentioned above.
Additionally, the central log ratio (*clr*) transformation was used to account for the compositional nature of microbial 16S data (*451*). Then, a PCA was used to observe and detect any bias, outlier or batch effect and look for patterns distinguishing males and females. The *sva R* package with its *ComBat()* function was used to correct a sampling batch effect. Finally, non-parametric Wilcoxon Rank Sum test was performed and significance level was set to $p < 0.05$ after correction for multiple testing with the BH procedure.

DNA methylation analysis on liver samples

DNA extraction and quality control

DNA was extracted using the the kit Qiamp DNA mini kit (Qiagen, Hilden, Germany), quality was assessed with the DNA 12000 kit from the Bioanalyzer 2100 Assay (Agilent, Santa Clara, CA). DNA was quantified using the fluorescent assay Qubit dsDNA kit measured with a Qubit 4.0 (Thermo Fisher Scientific,Waltham, US).

Library construction and data acquisition

Whole Genome Bisulfite Sequencing (WGBS) samples were prepared according to manufacturer instructions with the Accel-NGS Methyl-Seq DNA Library preparation (IDT, Coralville, US), and sequenced on Illumina NOVASEQ 6000 platform with a read length of 100bp PE, according to manufacturer instructions (Illumina, San Diego, US).

Data processing and statistical analysis

Paired end WGBS data was aligned and processed using the DKFZ OTP WGBS workflow version 1.2.73-202 against the GRCm38mm10 reference genome (*699*). Downstream analysis was performed with a custom *Snakemake* workflow, using *R* 4.1 with *Bioconductor* 3.14 (*700*, *701*). Methylation was called using *methyldackel* version 0.5.1 and the data was aggregated using the *methrix* package version 1.8.0(*702*). Pairwise differentially methylated regions were called using *DMRseq* version 1.14.0 (*703*), with following settings: *minInSpan*=11, *bpSpan*=500, *maxGapSmooth*=2000, *maxgap*=200, *cutoff*=0.05, *minRegions*=3.

DNA methylation functional analysis

Mapping to the reference genome found around 20k differentially methylated regions. The 20k DMRs were obtained by calculating the levels of methylation between each group in all the comparisons described previously. DMRs that overlapped each other were merged into one region and functional annotations were added. Around 11k DMRs overlapped and the mean methylation for each individual DMR was calculated per sample to be used in the correlational and integrative analysis.

Unsupervised integration of multi-omics data with MOFA2

To integrate the different omics data generated in the mice model, a Bayesian factor analysis linear model was used (*468*). Multi-omics factor analysis 2 (*MOFA2*) *R* package (v 1.8.0) provided higher interpretability compared to other non-linear models, allowing direct interpretation of the learned latent factors. The *MOFA* framework can be seen as an extension of *PCA* but in a multi-omic space, then learned latent factors can be seen as equivalent to principal components in a multi-omic space. *MOFA2* is a vertical integration method where measures of different omics have been taken on the same samples. One main *MOFA* model was constructed for all the liver measurements which included RNA-seq data, DNA methylation data, histone PTMs data, metals measurements, targeted metabolomics data and untargeted metabolomics data, as a main systems readout.

The input data for the *MOFA* model was: all RNA-seq differentially expressed genes (DEGs) in any of the comparisons (n=9942) properly transformed with the *vst()* function from *DESeq2*; the top 10k differentially methylated regions (DMRs) that were obtained by filtering top weighted IDs from the first component in a PCA analysis and transformed to M values (see formula below); all the untargeted metabolites that were differentially abundant between conditions with values transformed with the *varianceStabilizingTransformation*() function after size factor and dispersion were estimated; all targeted metabolites being also transformed with the *varianceStabilizingTransformation*() function: all histone

PTMs relative abundances transformed to M values as for the DMRs; and all the metals concentrations properly transformed with the *varianceStabilizingTransformation*() function after size factor and dispersion estimation.
All datasets, and previous *vst()*, *clr()* or M transformations were first inspected individually with the *mixOmics* package and using its *pca()*, *plotIndiv()* and *plotLoadings()* function to inspect the data and spot any batch effect or technical variability that needed to be addressed. The M values were calculated with the following formula:

$$Mi = log2((max(yi, methy, 0) + 0.99)/(max(yi, unmethy, 0) + 0.01))$$

After running the *run_mofa()* function, the weights for the first 3 factors were collected with the function *get_weights*() specifying the *scale=TRUE* parameter to have weights scaled from -1 to 1, within the same factor. The omics individual or collectively weights were used to filter, rank and visualize the sign in functional enrichment analysis such as GSEA or ORA.

Integrative correlational analysis

The integrative correlation analysis aimed to use the Pearson correlation *cor()*, to calculate the correlations between pairs of variables in each data modality. This approach has been used to fine-grain the results obtained from *MOFA* and leverage pairwise comparison between datasets, especially between 16S data (not included in the *MOFA* model) and the other data modalities. When the 16S data was used for correlation, only the corresponding samples in the other dataset were used. The transformed values used to run *MOFA* were used as input for the correlations. For the 16S data, all taxonomic annotations from phylum to species in colon and cecum with relative abundances values, were transformed with the central log ratio transformation *clr()* from the *compositions* (2.0-6) *R* package, to account for its compositional nature. All taxonomic levels present in one matrix (i.e. colon) were merged in a single dataframe to usi in the correlations.

The following datasets were correlated variable to variable for the same set of samples: liver DMRs-liver RNA, liver DMRs-liver untargeted metabolomics, liver RNA-liver untargeted metabolomics, liver DMR-liver targeted metabolites, liver DMRs-colon taxons, liver RNAs-colon taxons and colon taxons-liver untargeted metabolites. With the putative metabolites IDs, the annotated genes near a DMRs, or with the RNA gene IDs of top correlated features (abs(r) > 0.85), an overrepresentation test was performed with the *compareCluster()* function from *clusterProfiler* or *CalculateHyperScore()* from *metaboAnalystR*.

General discussion

Microbes have been essential in the evolution of all life forms present on Earth. They shape how plants extract nutrients from the soil or how humans mature their immune system. Microbes embarked in a myriad of symbiotic relationships with every living organism on this planet, colonizing almost all locations in their hosts. Because of this intimate relationship between the host and its microbiota, they form a superorganism called holobiont, functioning as a single biological, ecological and evolutionary unit. Meaning that their genetic background (including sex), their metabolism and their response to the environment are coordinated and complementary in health, or disturbed during disease.

When scientists try to study a human disease or biological characteristic using model organisms, for example mice, we tend to analyze molecular layers on the host and excluding the microbiota to explain multifactorial disease or complex biological systems. Losing the perspective of mice being a holobiont, where the delicate balance between the host and its microbiota may be one of the underlying reasons for staying healthy or suffering a disease.

I have developed my PhD research to understand better how the holobiont (host + microbiota) cometabolism interact and can modulate the host epigenome and gene transcription, promoting health or increasing the risk of suffering a disease. To do so, in Chapter 1, I contributed to better understand the nutrient-microbiota host metabolism-epigenetic axis of communication, by proposing the biological framework where to contextualize this axis. In Chapter 2, I contributed to better characterize key epigenetic metabolite players in this axis by creating two complementary metabolomics methods, one to interrogate the abundances of over 30 epigenetically relevant metabolites, and the other to measure the fate of labeled carbon C^{13} atoms in those epigenetic metabolites in a trace labeling experiment. And in Chapter 3, I contributed to understand better how the microbiota contributes to host epigenetic marks in a microbe and sex-dependent manner, by applying the novel metabolomics method to a germ-free mice model, together with measuring other epigenetically relevant molecular layers such as DNA methylation, RNA expression or histone PTMs.

Epigenetically relevant metabolites are endogenously or exogenously produced, and function as substrates, products, activators or inhibitors of epigenetic reactions. The addition or removal of these epigenetic marks on the DNA, RNA or proteins, such as methylation or acetylation, are known to switch on, and off host genes in a context-dependent manner (*331*). In addition, some epigenetic marks are reversible (*704*, *705*), opening the possibility of being modulated. Until now, no standardized metabolomics method existed to comprehensively interrogate and measure the abundances of many epigenetically relevant metabolites, at the same time. The novelty of the method I developed relies on the ability to measure many host endogenous epigenetically relevant metabolites, as well as being able to measure short chain fatty acids (SCFAs), using a single and simple metabolite extraction. The extraction solution consists of two immiscible solutions, that are analyzed on a LC-MS/MS or a GC-MS platform, to measure epigenetically relevant metabolites and SCFAs, respectively.

Historically, SCFAs have been analyzed by mass spectrometry approaches using a chemical derivatization step which is time consuming and introduces noise in the signal. The method presented here uses a chemical derivatization-free protocol to study them, together with other epigenetically

relevant metabolites. SCFAs are mainly produced in the intestinal tract by the resident microbiota after the fermentation of indigestible dietary fiber. Then, SCFAs can be absorbed to the bloodstream and virtually reach any cell in the body. SCFAs are also known to exert functions as histone deacetylase inhibitors (*88*), or serve as substrate to modify histones (*89*). Similarly to SCFAs, all metabolites included in our method have a direct or indirect epigenetic function for the host, and almost all of them can be potentially modulated by the microbiota.

The second targeted metabolomics approach presented in Chapter 2, allowed us to interrogate which are the pathways involved in building the moieties and biosynthesize three epigenetically relevant metabolites: S-adenosylmethionine, acetyl-CoA and UDP-N-acetylglucosamine, the main methyl, acetyl and glycosyl donors in epigenetic modifications, respectively. By trassing stable isotope labeled C^{13} atoms from an experiment where samples have been exposed to a labeled metabolite (i.e. glucose or glutamine), the method can return positional isotopomer information, rather than only indicating the number of incorporated labeled atoms, which is the main output in standard MS1 labeling experiments. The method allows to know the position where labeled atoms are incorporated, and estimate the rate or flux of the specific pathways involved in their biosynthesis.

These two novel targeted metabolomics methods are not without limitations. Some metabolites are susceptible to being rapidly oxidized, such as homocysteine or dihydrofolate with known epigenetic function, and are hard to preserve with the proposed extraction. Those metabolites would need the addition of specific antioxidant compounds to avoid their degradation, but compromising the measurement of the other metabolites. Another limitation in the first method is that, although no chemical derivatization is needed to analyze SCFAs, they still need to be analyzed with a GC-MS platform, while the water fraction with bigger metabolites, is analyzed by an LC-MS/MS platform. In the second method, one limitation is the contribution of naturally occurring labeled atoms in our signal. One needs to take into account the natural isotopic distribution and subtract it from the signal, to avoid overestimating the amount of incorporated labeled atoms.

In a biological context, it is clear that bidirectional metabolic communication between the host and its microbiota exist, for example via SCFAs (*90*). But microbiota can produce and consume important host endogenous metabolites such as succinate (*27*), opening the door to imagine that other endogenous epigenetically relevant metabolites are produced, consumed or influenced by the gut microbiota. The targeted metabolomics method presented in Chapter 2, will allow researchers to measure a variety of epigenetically relevant metabolites at the same time in their model system. Gaining substantial insights in how the metabolism (host + microbiota) is affecting the pool of these metabolites in their samples, and potentially affecting epigenetic marks. In this regard, very few research groups study how the host microbiota affects the host epigenetic landscape and try to directly address some of the following questions: can the gut microbiota modulate the epigenetic marks in the host in a distant organ, for example the liver, and turn on and off, host genes there? Which epigenetic marks are affected, marks on DNA, RNA or on proteins? Which taxons contribute the most to these mechanisms? There are specific bacteria that target specific genes or functions? And the "language" of communication between the host and its microbiota, is it only metabolites? or do they also exchange small genetic fragments? A different microbiota composition between males and females, affects the microbiota-host communication?

In Chapter 3, we sought to partially answer some of these questions by characterizing a germ-free mice model. Investigating the effect of the microbiota and sex of the host, in host epigenetic marks, gene transcription, abundances of epigenetically relevant metabolites or other metabolites. We took a multi-omics approach and measured several data modalities in the same host mice and its microbiota. The biological rationale was to leverage the biological interdependency of the different molecular layers measured in the holobiont, to obtain an holistic view of the system, and be able to spot interactions between the different molecular layers in the same individual.

After analyzing the data we have made several important observations that connected the microbiome to the epigenetic marks in the host. The first and more striking observation is that DNA methylation levels depend on the presence or absence of microbiota in a sex-dependent manner. Having conventional males a global DNA hypomethylation, and germ-free females a global DNA hypermethylation, while germ-free males and conventional females have similar levels. The second related observation is the additivity effect of sex and microbiota exerted in the DNA methylation levels, where conventional animales (females and males), had lower levels of DNA methylation, compared to their germ-free counterparts. The third related observation is that males are hypomethylated respect females in both conditions, conventional and germ-free, likely driven by testosterone levels. We further observed that conventional males had significantly lower levels of methylation in intronic regions compared to their exonic and untranslated regions. But when the same comparison was investigated in the other three groups, no significant differences were observed between the levels of methylation in intronic regions, their exons or untranslated regions. Suggesting microbiota in males has a role in controlling methylation levels in specific genomic locations. Additionally, the global hypomethylation patterns in conventional males DNA, strongly correlated to a major number of expressed genes in this group.

The correlation analysis between the taxons present in the colon and the levels of DNA methylation or gene expression in the liver, show strong correlations between the levels of low abundant taxons like *Melainabacteria* and the expression of few genes related to mitochondria, signal transduction and lipid metabolism. And strong correlations with higher abundant taxons, like the *Ruminococcaceae* family members, and the levels of methylation and gene expression of many genes, correlated to steroid and fatty acid metabolism genes. Remarkably, the methylation levels in 4 DMR genes were positively correlated to the levels of *Ruminococcaceae*, being these genes higher methylated when the *Ruminococcaceae* members are more abundant, or vice versa, while the rest (> 100 DMR genes) correlated negatively. The positive correlated DMR genes are key genes in the degradation of testosterone, which need to be in strict control in a sex-dependent manner, indicating a potential role of the *Ruminococcaceae* family in controlling the expression of testosterone degrading genes through its hypermethylation in conventional males.

These remarkable findings allow us to speculate how steroid metabolism and mitochondria could be one of the targeted metabolism and organelles for the microbiota metabolism. Where the microbiota would produce metabolites that specifically modulate mitochondria dynamics and steroid metabolism, through the epigenetic control of few genes, with a global DNA methylation effect. Reduced glutathione can be produced by the host and some species in the microbiota, offering another possible

link of how the microbiota might be affecting the mitochondria (*706*). In a similar regard, mitochondria-microbiota communication has been suggested to balance, sense and regulate oxidative stress and lipid metabolism in the host, or modulating the microbiota composition through excreted reactive oxygen species in the gut (*201*). And targeting this communication between the host mitochondria and its microbiota, has been proposed to be a possible therapeutic intervention framework to treat metabolic disease and age related diseases (*707*).

Going beyond a metabolic communication between the host and its microbiota, it has been shown how host enterocyte cells in the intestine release micro-RNAs that target controlling and modulating the gut microbiota composition (*346*). It has also been proposed that the plethora of genomic antisense molecules produced by the microbiota are merely transcriptional noise (*708*), but I like to think that it is another layer of host-microbiota communication that has not been well studied. Where the microbiome could also communicate with their host genome, modulating host genetic and metabolic activity, sending for example, small pieces of antisense DNA through the bloodstream targeting specific host transcripts. Then, these small pieces of DNA or RNA, could interact with the host genome/transcriptome in a distant organ, and change its expression patterns. This idea is little explored and warrants further investigation.

Altogether, we contributed with the creation of a targeted metabolomics method allowing the study of the host-microbiota communication through epigenetically relevant metabolites, and their effects on the host epigenome. The method has already been applied to study epigenetic-metabolism events in animal models of metabolic disease and cell culture experiments. We have also contributed to better understand how the microbiota can affect and correlate to epigenetic marks in a sex-dependent manner, stressing the importance of including microbiota and sex, as two important experimental factors in the design and execution of future research projects aiming to conduct research in the interphase between microbiome, metabolism and epigenetics.

Concluding remarks

Here I list the main conclusions of this thesis, related to the main objectives of the PhD:

- The holobiont, brings us an excellent biological, ecological and evolutionary framework to integrate and understand the communications between the host and its microbiota through the nutrient-microbe metabolism-host epigenetic axis of communication.

- The targeted metabolomics method we have created, allows us to interrogate epigenetically relevant metabolites, including microbiota produced short chain fatty acids, S-adenosylmethionine or acetyl-CoA, among other epigenetic metabolites, in a simple and easy to implement experimental procedure.

- The multi-omics study of the germ-free and conventional mice model has revealed how the microbiota composition and sex influence the host DNA methylation and transcriptomic patterns in a sex and microbiota additivity manner, highlighting steroid metabolism

Bibliography

1. R. Sender, S. Fuchs, R. Milo, Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol.* **14**, e1002533 (2016).

2. J. Qin, R. Li, J. Raes, M. Arumugam, K. S. Burgdorf, C. Manichanh, T. Nielsen, N. Pons, F. Levenez, T. Yamada, D. R. Mende, J. Li, J. Xu, S. Li, D. Li, J. Cao, B. Wang, H. Liang, H. Zheng, Y. Xie, J. Tap, P. Lepage, M. Bertalan, J.-M. Batto, T. Hansen, D. Le Paslier, A. Linneberg, H. B. Nielsen, E. Pelletier, P. Renault, T. Sicheritz-Ponten, K. Turner, H. Zhu, C. Yu, S. Li, M. Jian, Y. Zhou, Y. Li, X. Zhang, S. Li, N. Qin, H. Yang, J. Wang, S. Brunak, J. Doré, F. Guarner, K. Kristiansen, O. Pedersen, J. Parkhill, J. Weissenbach, MetaHIT Consortium, P. Bork, S. D. Ehrlich, J. Wang, A human gut microbial gene catalogue established by metagenomic sequencing. *Nature.* **464**, 59–65 (2010).

3. N. Kitadai, S. Maruyama, Origins of building blocks of life: A review. *Geosci. Front.* **9**, 1117–1153 (2018).

4. B. Liu, C. G. Pappas, J. Ottelé, G. Schaeffer, C. Jurissek, P. F. Pieters, M. Altay, I. Marić, M. C. A. Stuart, S. Otto, Spontaneous Emergence of Self-Replicating Molecules Containing Nucleobases and Amino Acids. *J. Am. Chem. Soc.* **142**, 4184–4192 (2020).

5. J. D. Sutherland, Opinion: Studies on the origin of life — the end of the beginning. *Nat. Rev. Chem.* **1**, 0012 (2017).

6. N. Glansdorff, Y. Xu, B. Labedan, The last universal common ancestor: emergence, constitution and genetic legacy of an elusive forerunner. *Biol. Direct.* **3**, 29 (2008).

7. E. L. Madsen, Microorganisms and their roles in fundamental biogeochemical cycles. *Curr. Opin. Biotechnol.* **22**, 456–464 (2011).

8. P. G. Falkowski, T. Fenchel, E. F. Delong, The microbial engines that drive Earth's biogeochemical cycles. *Science.* **320**, 1034–1039 (2008).

9. R. E. Ley, M. Hamady, C. Lozupone, P. J. Turnbaugh, R. R. Ramey, J. S. Bircher, M. L. Schlegel, T. A. Tucker, M. D. Schrenzel, R. Knight, J. I. Gordon, Evolution of mammals and their gut microbes. *Science.* **320**, 1647–1651 (2008).

10. L. Sagan, On the origin of mitosing cells. *J. Theor. Biol.* **14**, 255–274 (1967).

11. K. Zaremba-Niedzwiedzka, E. F. Caceres, J. H. Saw, D. Bäckström, L. Juzokaite, E. Vancaester, K. W. Seitz, K. Anantharaman, P. Starnawski, K. U. Kjeldsen, M. B. Stott, T. Nunoura, J. F. Banfield, A. Schramm, B. J. Baker, A. Spang, T. J. G. Ettema, Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature.* **541**, 353–358 (2017).

12. L. A. Hug, B. J. Baker, K. Anantharaman, C. T. Brown, A. J. Probst, C. J. Castelle, C. N. Butterfield, A. W. Hernsdorf, Y. Amano, K. Ise, Y. Suzuki, N. Dudek, D. A. Relman, K. M. Finstad, R. Amundson, B. C. Thomas, J. F. Banfield, A new view of the tree of life. *Nat Microbiol.* **1**, 16048 (2016).

13. T. W. Lyons, C. T. Reinhard, N. J. Planavsky, The rise of oxygen in Earth's early ocean and atmosphere. *Nature.* **506**, 307–315 (2014).

14. M. F. Hohmann-Marriott, R. E. Blankenship, Evolution of photosynthesis. *Annu. Rev. Plant Biol.* **62**, 515–548 (2011).

15. C. R. Woese, O. Kandler, M. L. Wheelis, Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 4576–4579 (1990).

16. C. R. Woese, G. E. Fox, Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5088–5090 (1977).

17. E. Rosenberg, I. Zilber-Rosenberg, Microbes Drive Evolution of Animals and Plants: the Hologenome Concept. *MBio.* **7**, e01395 (2016).

18. L. Margulis, R. Fester, Bellagio conference and book. Symbiosis as Source of Evolutionary Innovation: Speciation and Morphogenesis. Conference--June 25-30, 1989, Bellagio Conference Center, Italy. *Symbiosis.* **11**, 93–101 (1991).

19. W. Martin, K. Kowallik, Annotated English translation of Mereschkowsky's 1905 paper "Über Natur und Ursprung der Chromatophoren imPflanzenreiche." *Eur. J. Phycol.* **34**, 287–295 (1999).

20. J. Martijn, J. Vosseberg, L. Guy, P. Offre, T. J. G. Ettema, Deep mitochondrial origin outside the sampled alphaproteobacteria. *Nature.* **557**, 101–105 (2018).

21. H. Imachi, M. K. Nobu, N. Nakahara, Y. Morono, M. Ogawara, Y. Takaki, Y. Takano, K. Uematsu, T. Ikuta, M. Ito, Y. Matsui, M. Miyazaki, K. Murata, Y. Saito, S. Sakai, C. Song, E. Tasumi, Y. Yamanaka, T. Yamaguchi, Y. Kamagata, H. Tamaki, K. Takai, Isolation of an archaeon at the prokaryote-eukaryote interface. *Nature.* **577**, 519–525 (2020).

22. J. Miro-Blanch, O. Yanes, Epigenetic Regulation at the Interplay Between Gut Microbiota and Host Metabolism. *Front. Genet.* **10**, 638 (2019).

23. S. Katada, A. Imhof, P. Sassone-Corsi, Connecting threads: epigenetics and metabolism. *Cell.* **148**, 24–28 (2012).

24. J. L. Sonnenburg, F. Bäckhed, Diet–microbiota interactions as moderators of human metabolism. *Nature.* **535**, 56–64 (2016).

25. K. A. Krautkramer, J. H. Kreznar, K. A. Romano, E. I. Vivas, G. A. Barrett-Wilt, M. E. Rabaglia, M. P. Keller, A. D. Attie, F. E. Rey, J. M. Denu, Diet-Microbiota Interactions Mediate Global Epigenetic Programming in Multiple Host Tissues. *Mol. Cell.* **64**, 982–992 (2016).

26. J. G. LeBlanc, C. Milani, G. S. de Giori, F. Sesma, D. van Sinderen, M. Ventura, Bacteria as vitamin suppliers to their host: a gut microbiota perspective. *Curr. Opin. Biotechnol.* **24**, 160–168 (2013).

27. S. Fernández-Veledo, J. Vendrell, Gut microbiota-derived succinate: Friend or foe in human metabolic diseases? *Rev. Endocr. Metab. Disord.* **20**, 439–447 (2019).

28. K. P. Scott, S. H. Duncan, H. J. Flint, Dietary fibre and the gut microbiota. *Nutr. Bull.* **33**, 201–211 (2008).

29. I. Rowland, G. Gibson, A. Heinken, K. Scott, J. Swann, I. Thiele, K. Tuohy, Gut microbiota functions: metabolism of nutrients and other food components. *Eur. J. Nutr.* **57**, 1–24 (2018).

30. H. Neuman, J. W. Debelius, R. Knight, O. Koren, Microbial endocrinology: the interplay between

the microbiota and the endocrine system. *FEMS Microbiol. Rev.* **39**, 509–521 (2015).

31. P. Strandwitz, Neurotransmitter modulation by the gut microbiota. *Brain Res.* **1693**, 128–133 (2018).

32. H.-J. Wu, E. Wu, The role of gut microbiota in immune homeostasis and autoimmunity. *Gut Microbes*. **3**, 4–14 (2012).

33. Q. Ma, C. Xing, W. Long, H. Y. Wang, Q. Liu, R.-F. Wang, Impact of microbiota on central nervous system and neurological diseases: the gut-brain axis. *J. Neuroinflammation*. **16**, 53 (2019).

34. R. A. Quinn, A. V. Melnik, A. Vrbanac, T. Fu, K. A. Patras, M. P. Christy, Z. Bodai, P. Belda-Ferre, A. Tripathi, L. K. Chung, M. Downes, R. D. Welch, M. Quinn, G. Humphrey, M. Panitchpakdi, K. C. Weldon, A. Aksenov, R. da Silva, J. Avila-Pacheco, C. Clish, S. Bae, H. Mallick, E. A. Franzosa, J. Lloyd-Price, R. Bussell, T. Thron, A. T. Nelson, M. Wang, E. Leszczynski, F. Vargas, J. M. Gauglitz, M. J. Meehan, E. Gentry, T. D. Arthur, A. C. Komor, O. Poulsen, B. S. Boland, J. T. Chang, W. J. Sandborn, M. Lim, N. Garg, J. C. Lumeng, R. J. Xavier, B. I. Kazmierczak, R. Jain, M. Egan, K. E. Rhee, D. Ferguson, M. Raffatellu, H. Vlamakis, G. G. Haddad, D. Siegel, C. Huttenhower, S. K. Mazmanian, R. M. Evans, V. Nizet, R. Knight, P. C. Dorrestein, Global chemical effects of the microbiome include new bile-acid conjugations. *Nature*. **579**, 123–129 (2020).

35. J. F. Cryan, T. G. Dinan, Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour. *Nat. Rev. Neurosci.* **13**, 701–712 (2012).

36. H. Han, B. Yi, R. Zhong, M. Wang, S. Zhang, J. Ma, Y. Yin, J. Yin, L. Chen, H. Zhang, From gut microbiota to host appetite: gut microbiota-derived metabolites as key regulators. *Microbiome*. **9**, 162 (2021).

37. M. Gurung, Z. Li, H. You, R. Rodrigues, D. B. Jump, A. Morgun, N. Shulzhenko, Role of gut microbiota in type 2 diabetes pathophysiology. *EBioMedicine*. **51**, 102590 (2020).

38. A. A. Kolodziejczyk, D. Zheng, E. Elinav, Diet-microbiota interactions and personalized nutrition. *Nat. Rev. Microbiol.* **17**, 742–753 (2019).

39. N. Mach, D. Fuster-Botella, Endurance exercise and gut microbiota: A review. *J Sport Health Sci.* **6**, 179–197 (2017).

40. C. A. Thaiss, M. Levy, T. Korem, L. Dohnalová, H. Shapiro, D. A. Jaitin, E. David, D. R. Winter, M. Gury-BenAri, E. Tatirovsky, T. Tuganbaev, S. Federici, N. Zmora, D. Zeevi, M. Dori-Bachash, M. Pevsner-Fischer, E. Kartvelishvily, A. Brandis, A. Harmelin, O. Shibolet, Z. Halpern, K. Honda, I. Amit, E. Segal, E. Elinav, Microbiota Diurnal Rhythmicity Programs Host Transcriptome Oscillations. *Cell.* **167**, 1495–1510.e12 (2016).

41. D. L. Nelson, A. L. Lehninger, M. M. Cox, *Lehninger Principles of Biochemistry* (W. H. Freeman, 2008).

42. C. M. Metallo, M. G. Vander Heiden, Understanding metabolic regulation and its influence on cell physiology. *Mol. Cell.* **49**, 388–398 (2013).

43. A. A. Gibb, B. G. Hill, Metabolic Coordination of Physiological and Pathological Cardiac Remodeling. *Circ. Res.* **123**, 107–128 (2018).

44. C. M. O'Brien, B. C. Mulukutla, D. G. Mashek, W.-S. Hu, Regulation of Metabolic Homeostasis in

Cell Culture Bioprocesses. *Trends Biotechnol.* **38**, 1113–1127 (2020).

45. J. R. Brestoff, D. Artis, Immune regulation of metabolic homeostasis in health and disease. *Cell.* **161**, 146–160 (2015).

46. J. Smeitink, L. van den Heuvel, S. DiMauro, The genetics and pathology of oxidative phosphorylation. *Nat. Rev. Genet.* **2**, 342–352 (2001).

47. C. E. Weinberg, Z. Weinberg, C. Hammann, Novel ribozymes: discovery, catalytic mechanisms, and the quest to understand biological function. *Nucleic Acids Res.* **47**, 9480–9494 (2019).

48. E. Holmes, I. D. Wilson, J. K. Nicholson, Metabolic phenotyping in health and disease. *Cell.* **134**, 714–717 (2008).

49. J. L. DeRisi, V. R. Iyer, P. O. Brown, Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science.* **278**, 680–686 (1997).

50. J. A. van der Knaap, C. P. Verrijzer, Undercover: gene control by metabolites and metabolic enzymes. *Genes Dev.* **30**, 2345–2369 (2016).

51. X. Li, G. Egervari, Y. Wang, S. L. Berger, Z. Lu, Regulation of chromatin and gene expression by metabolic enzymes and metabolites. *Nat. Rev. Mol. Cell Biol.* **19**, 563–578 (2018).

52. J. Zhu, C. B. Thompson, Metabolic regulation of cell growth and proliferation. *Nat. Rev. Mol. Cell Biol.* **20**, 436–450 (2019).

53. M. Lempp, N. Farke, M. Kuntz, S. A. Freibert, R. Lill, H. Link, Systematic identification of metabolites controlling gene expression in E. coli. *Nat. Commun.* **10**, 4463 (2019).

54. K. Wang, W. Zhou, P. Meng, P. Wang, C. Zhou, Y. Yao, S. Wu, Y. Wang, J. Zhao, D. Zou, G. Jin, Immune-related somatic mutation genes are enriched in PDACs with diabetes. *Transl. Oncol.* **12**, 1147–1154 (2019).

55. S. Zhao, W. Xu, W. Jiang, W. Yu, Y. Lin, T. Zhang, J. Yao, L. Zhou, Y. Zeng, H. Li, Y. Li, J. Shi, W. An, S. M. Hancock, F. He, L. Qin, J. Chin, P. Yang, X. Chen, Q. Lei, Y. Xiong, K.-L. Guan, Regulation of cellular metabolism by protein lysine acetylation. *Science.* **327**, 1000–1004 (2010).

56. S. M. Sanderson, X. Gao, Z. Dai, J. W. Locasale, Methionine metabolism in health and cancer: a nexus of diet and precision medicine. *Nat. Rev. Cancer.* **19**, 625–637 (2019).

57. G. S. Hotamisligil, Inflammation and metabolic disorders. *Nature.* **444**, 860–867 (2006).

58. P. C. Calder, N. Ahluwalia, F. Brouns, T. Buetler, K. Clement, K. Cunningham, K. Esposito, L. S. Jönsson, H. Kolb, M. Lansink, A. Marcos, A. Margioris, N. Matusheski, H. Nordmann, J. O'Brien, G. Pugliese, S. Rizkalla, C. Schalkwijk, J. Tuomilehto, J. Wärnberg, B. Watzl, B. M. Winklhofer-Roob, Dietary factors and low-grade inflammation in relation to overweight and obesity. *Br. J. Nutr.* **106 Suppl 3**, S5–78 (2011).

59. M.-E. Patti, S. Corvera, The role of mitochondria in the pathogenesis of type 2 diabetes. *Endocr. Rev.* **31**, 364–395 (2010).

60. A. Gonzalez-Franquesa, P. Gama-Perez, M. Kulis, K. Szczepanowska, N. Dahdah, S. Moreno-Gomez, A. Latorre-Pellicer, R. Fernández-Ruiz, A. Aguilar-Mogas, A. Hoffman, E.

Monelli, S. Samino, J. Miró-Blanch, G. Oemer, X. Duran, E. Sanchez-Rebordelo, M. Schneeberger, M. Obach, J. Montane, G. Castellano, V. Chapaprieta, W. Sun, L. Navarro, I. Prieto, C. Castaño, A. Novials, R. Gomis, M. Monsalve, M. Claret, M. Graupera, G. Soria, C. Wolfrum, J. Vendrell, S. Fernández-Veledo, J. A. Enríquez, A. Carracedo, J. C. Perales, R. Nogueiras, L. Herrero, A. Trifunovic, M. A. Keller, O. Yanes, M. Sales-Pardo, R. Guimerà, M. Blüher, J. I. Martín-Subero, P. M. Garcia-Roves, Remission of obesity and insulin resistance is not sufficient to restore mitochondrial homeostasis in visceral adipose tissue. *Redox Biol.* **54**, 102353 (2022).

61. P. Manna, S. K. Jain, Obesity, Oxidative Stress, Adipose Tissue Dysfunction, and the Associated Health Risks: Causes and Therapeutic Strategies. *Metab. Syndr. Relat. Disord.* **13**, 423–444 (2015).

62. W. Wang, F. Zhao, X. Ma, G. Perry, X. Zhu, Mitochondria dysfunction in the pathogenesis of Alzheimer's disease: recent advances. *Mol. Neurodegener.* **15**, 30 (2020).

63. A. H. de Mello, A. B. Costa, J. D. G. Engel, G. T. Rezin, Mitochondrial dysfunction in obesity. *Life Sci.* **192**, 26–32 (2018).

64. D. E. Meyers, H. I. Basha, M. K. Koenig, Mitochondrial cardiomyopathy: pathophysiology, diagnosis, and management. *Tex. Heart Inst. J.* **40**, 385–394 (2013).

65. S. Vyas, E. Zaganjor, M. C. Haigis, Mitochondria and Cancer. *Cell.* **166**, 555–566 (2016).

66. P. D. Cani, N. M. Delzenne, The gut microbiome as therapeutic target. *Pharmacol. Ther.* **130**, 202–212 (2011).

67. I. C. L. van den Munckhof, A. Kurilshikov, R. Ter Horst, N. P. Riksen, L. A. B. Joosten, A. Zhernakova, J. Fu, S. T. Keating, M. G. Netea, J. de Graaf, J. H. W. Rutten, Role of gut microbiota in chronic low-grade inflammation as potential driver for atherosclerotic cardiovascular disease: a systematic review of human studies. *Obes. Rev.* **19**, 1719–1734 (2018).

68. K. Mokkala, N. Houttu, E. Koivuniemi, N. Sørensen, H. B. Nielsen, K. Laitinen, GlycA, a novel marker for low grade inflammation, reflects gut microbiome diversity and is more accurate than high sensitive CRP in reflecting metabolomic profile. *Metabolomics*. **16**, 76 (2020).

69. R. E. Ley, F. Bäckhed, P. Turnbaugh, C. A. Lozupone, R. D. Knight, J. I. Gordon, Obesity alters gut microbial ecology. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 11070–11075 (2005).

70. L. Wen, R. E. Ley, P. Y. Volchkov, P. B. Stranges, L. Avanesyan, A. C. Stonebraker, C. Hu, F. S. Wong, G. L. Szot, J. A. Bluestone, J. I. Gordon, A. V. Chervonsky, Innate immunity and intestinal microbiota in the development of Type 1 diabetes. *Nature*. **455**, 1109–1113 (2008).

71. M. Y. Lim, H. J. You, H. S. Yoon, B. Kwon, J. Y. Lee, S. Lee, Y.-M. Song, K. Lee, J. Sung, G. Ko, The effect of heritability and host genetics on the gut microbiota and metabolic syndrome. *Gut.* **66**, 1031–1038 (2017).

72. S. Yachida, S. Mizutani, H. Shiroma, S. Shiba, T. Nakajima, T. Sakamoto, H. Watanabe, K. Masuda, Y. Nishimoto, M. Kubo, F. Hosoda, H. Rokutan, M. Matsumoto, H. Takamaru, M. Yamada, T. Matsuda, M. Iwasaki, T. Yamaji, T. Yachida, T. Soga, K. Kurokawa, A. Toyoda, Y. Ogura, T. Hayashi, M. Hatakeyama, H. Nakagama, Y. Saito, S. Fukuda, T. Shibata, T. Yamada, Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nat. Med.* **25**, 968–976 (2019).

73. B. Routy, E. Le Chatelier, L. Derosa, C. P. M. Duong, M. T. Alou, R. Daillère, A. Fluckiger, M.

Messaoudene, C. Rauber, M. P. Roberti, M. Fidelle, C. Flament, V. Poirier-Colame, P. Opolon, C. Klein, K. Iribarren, L. Mondragón, N. Jacquelot, B. Qu, G. Ferrere, C. Clémenson, L. Mezquita, J. R. Masip, C. Naltet, S. Brosseau, C. Kaderbhai, C. Richard, H. Rizvi, F. Levenez, N. Galleron, B. Quinquis, N. Pons, B. Ryffel, V. Minard-Colin, P. Gonin, J.-C. Soria, E. Deutsch, Y. Loriot, F. Ghiringhelli, G. Zalcman, F. Goldwasser, B. Escudier, M. D. Hellmann, A. Eggermont, D. Raoult, L. Albiges, G. Kroemer, L. Zitvogel, Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. *Science*. **359**, 91–97 (2018).

74. S. Fromentin, S. K. Forslund, K. Chechi, J. Aron-Wisnewsky, R. Chakaroun, T. Nielsen, V. Tremaroli, B. Ji, E. Prifti, A. Myridakis, J. Chilloux, P. Andrikopoulos, Y. Fan, M. T. Olanipekun, R. Alves, S. Adiouch, N. Bar, Y. Talmor-Barkan, E. Belda, R. Caesar, L. P. Coelho, G. Falony, S. Fellahi, P. Galan, N. Galleron, G. Helft, L. Hoyles, R. Isnard, E. Le Chatelier, H. Julienne, L. Olsson, H. K. Pedersen, N. Pons, B. Quinquis, C. Rouault, H. Roume, J.-E. Salem, T. S. B. Schmidt, S. Vieira-Silva, P. Li, M. Zimmermann-Kogadeeva, C. Lewinter, N. B. Søndertoft, T. H. Hansen, D. Gauguier, J. P. Gøtze, L. Køber, R. Kornowski, H. Vestergaard, T. Hansen, J.-D. Zucker, S. Hercberg, I. Letunic, F. Bäckhed, J.-M. Oppert, J. Nielsen, J. Raes, P. Bork, M. Stumvoll, E. Segal, K. Clément, M.-E. Dumas, S. D. Ehrlich, O. Pedersen, Microbiome and metabolome features of the cardiometabolic disease spectrum. *Nat. Med.* **28**, 303–314 (2022).

75. F. J. Ryan, A. M. Ahern, R. S. Fitzgerald, E. J. Laserna-Mendieta, E. M. Power, A. G. Clooney, K. W. O'Donoghue, P. J. McMurdie, S. Iwai, A. Crits-Christoph, D. Sheehan, C. Moran, B. Flemer, A. L. Zomer, A. Fanning, J. O'Callaghan, J. Walton, A. Temko, W. Stack, L. Jackson, S. A. Joyce, S. Melgar, T. Z. DeSantis, J. T. Bell, F. Shanahan, M. J. Claesson, Colonic microbiota is associated with inflammation and host epigenomic alterations in inflammatory bowel disease. *Nat. Commun.* **11**, 1512 (2020).

76. D. N. Jackson, A. L. Theiss, Gut bacteria signaling to mitochondria in intestinal inflammation and cancer. *Gut Microbes*. **11**, 285–304 (2020).

77. T. Yardeni, C. E. Tanes, K. Bittinger, L. M. Mattei, P. M. Schaefer, L. N. Singh, G. D. Wu, D. G. Murdock, D. C. Wallace, Host mitochondria influence gut microbiome diversity: A role for ROS. *Sci. Signal.* **12** (2019), doi:10.1126/scisignal.aaw3159.

78. J. A. Bravo, P. Forsythe, M. V. Chew, E. Escaravage, H. M. Savignac, T. G. Dinan, J. Bienenstock, J. F. Cryan, Ingestion of Lactobacillus strain regulates emotional behavior and central GABA receptor expression in a mouse via the vagus nerve. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 16050–16055 (2011).

79. W. G. Kaelin Jr, S. L. McKnight, Influence of metabolism on epigenetics and disease. *Cell.* **153**, 56–69 (2013).

80. D. Zhang, Z. Tang, H. Huang, G. Zhou, C. Cui, Y. Weng, W. Liu, S. Kim, S. Lee, M. Perez-Neut, J. Ding, D. Czyz, R. Hu, Z. Ye, M. He, Y. G. Zheng, H. A. Shuman, L. Dai, B. Ren, R. G. Roeder, L. Becker, Y. Zhao, Metabolic regulation of gene expression by histone lactylation. *Nature.* **574**, 575–580 (2019).

81. J. Smestad, L. Erber, Y. Chen, L. J. Maher 3rd, Chromatin Succinylation Correlates with Active Gene Expression and Is Perturbed by Defective TCA Cycle Metabolism. *iScience.* **2**, 63–75 (2018).

82. Y. Lu, Q. Xu, Y. Liu, Y. Yu, Z.-Y. Cheng, Y. Zhao, D.-X. Zhou, Dynamics and functional interplay of histone lysine butyrylation, crotonylation, and acetylation in rice under starvation and submergence. *Genome Biol.* **19**, 144 (2018).

83. H. Shi, J. Wei, C. He, Where, When, and How: Context-Dependent Functions of RNA Methylation Writers, Readers, and Erasers. *Mol. Cell.* **74**, 640–650 (2019).

84. W. R. Wikoff, A. T. Anfora, J. Liu, P. G. Schultz, S. A. Lesley, E. C. Peters, G. Siuzdak, Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 3698–3703 (2009).

85. R. Corrêa-Oliveira, J. L. Fachi, A. Vieira, F. T. Sato, M. A. R. Vinolo, Regulation of immune cell function by short-chain fatty acids. *Clin Transl Immunology*. **5**, e73 (2016).

86. G. Sharon, T. R. Sampson, D. H. Geschwind, S. K. Mazmanian, The Central Nervous System and the Gut Microbiome. *Cell.* **167**, 915–932 (2016).

87. H. Liu, J. Wang, T. He, S. Becker, G. Zhang, D. Li, X. Ma, Butyrate: A Double-Edged Sword for Health? *Adv. Nutr.* **9**, 21–29 (2018).

88. J. R. Davie, Inhibition of histone deacetylase activity by butyrate. *J. Nutr.* **133**, 2485S–2493S (2003).

89. G. den Besten, K. van Eunen, A. K. Groen, K. Venema, D.-J. Reijngoud, B. M. Bakker, The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. *J. Lipid Res.* **54**, 2325–2340 (2013).

90. Y. P. Silva, A. Bernardi, R. L. Frozza, The Role of Short-Chain Fatty Acids From Gut Microbiota in Gut-Brain Communication. *Front. Endocrinol.* . **11**, 25 (2020).

91. G. den Besten, A. Bleeker, A. Gerding, K. van Eunen, R. Havinga, T. H. van Dijk, M. H. Oosterveer, J. W. Jonker, A. K. Groen, D.-J. Reijngoud, B. M. Bakker, Short-Chain Fatty Acids Protect Against High-Fat Diet-Induced Obesity via a PPARγ-Dependent Switch From Lipogenesis to Fat Oxidation. *Diabetes*. **64**, 2398–2408 (2015).

92. P. V. Chang, L. Hao, S. Offermanns, R. Medzhitov, The microbial metabolite butyrate regulates intestinal macrophage function via histone deacetylase inhibition. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 2247–2252 (2014).

93. K. M. Maslowski, C. R. Mackay, Diet, gut microbiota and immune responses. *Nat. Immunol.* **12**, 5–9 (2011).

94. Y. Chen, R. Sprung, Y. Tang, H. Ball, B. Sangras, S. C. Kim, J. R. Falck, J. Peng, W. Gu, Y. Zhao, Lysine propionylation and butyrylation are novel post-translational modifications in histones. *Mol. Cell. Proteomics*. **6**, 812–819 (2007).

95. S. Fiorucci, M. Biagioli, A. Zampella, E. Distrutti, Bile Acids Activated Receptors Regulate Innate Immunity. *Front. Immunol.* **9**, 1853 (2018).

96. L. F. Huergo, R. Dixon, The Emergence of 2-Oxoglutarate as a Master Regulator Metabolite. *Microbiol. Mol. Biol. Rev.* **79**, 419–435 (2015).

97. X. Gao, S. M. Sanderson, Z. Dai, M. A. Reid, D. E. Cooper, M. Lu, J. P. Richie Jr, A. Ciccarella, A. Calcagnotto, P. G. Mikhael, S. J. Mentch, J. Liu, G. Ables, D. G. Kirsch, D. S. Hsu, S. N. Nichenametla, J. W. Locasale, Dietary methionine influences therapy in mouse cancer models and alters human metabolism. *Nature*. **572**, 397–401 (2019).

98. M. J. Hill, Intestinal flora and endogenous vitamin synthesis. *Eur. J. Cancer Prev.* **6 Suppl 1**, S43–5

(1997).

99. J. Mayneris-Perxachs, M. Cardellini, L. Hoyles, J. Latorre, F. Davato, J. M. Moreno-Navarrete, M. Arnoriaga-Rodríguez, M. Serino, J. Abbott, R. H. Barton, J. Puig, X. Fernández-Real, W. Ricart, C. Tomlinson, M. Woodbridge, P. Gentileschi, S. A. Butcher, E. Holmes, J. K. Nicholson, V. Pérez-Brocal, A. Moya, D. M. Clain, R. Burcelin, M.-E. Dumas, M. Federici, J.-M. Fernández-Real, Iron status influences non-alcoholic fatty liver disease in obesity through the gut microbiome. *Microbiome*. **9** (2021), doi:10.1186/s40168-021-01052-7.

100. L. Yurkovetskiy, M. Burrows, A. A. Khan, L. Graham, P. Volchkov, L. Becker, D. Antonopoulos, Y. Umesaki, A. V. Chervonsky, Gender bias in autoimmunity is influenced by microbiota. *Immunity*. **39**, 400–412 (2013).

101. J. G. M. Markle, D. N. Frank, S. Mortin-Toth, C. E. Robertson, L. M. Feazel, U. Rolle-Kampczyk, M. von Bergen, K. D. McCoy, A. J. Macpherson, J. S. Danska, Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity. *Science*. **339**, 1084–1088 (2013).

102. J. Mayneris-Perxachs, M. Arnoriaga-Rodríguez, D. Luque-Córdoba, F. Priego-Capote, V. Pérez-Brocal, A. Moya, A. Burokas, R. Maldonado, J.-M. Fernández-Real, Gut microbiota steroid sexual dimorphism and its impact on gonadal steroids: influences of obesity and menopausal status. *Microbiome*. **8**, 136 (2020).

103. J. F. Cryan, K. J. O'Riordan, C. S. M. Cowan, K. V. Sandhu, T. F. S. Bastiaanssen, M. Boehme, M. G. Codagnone, S. Cussotto, C. Fulling, A. V. Golubeva, K. E. Guzzetta, M. Jaggar, C. M. Long-Smith, J. M. Lyte, J. A. Martin, A. Molinero-Perez, G. Moloney, E. Morelli, E. Morillas, R. O'Connor, J. S. Cruz-Pereira, V. L. Peterson, K. Rea, N. L. Ritz, E. Sherwin, S. Spichak, E. M. Teichman, M. van de Wouw, A. P. Ventura-Silva, S. E. Wallace-Fitzsimons, N. Hyland, G. Clarke, T. G. Dinan, The Microbiota-Gut-Brain Axis. *Physiol. Rev.* **99**, 1877–2013 (2019).

104. M. Valles-Colomer, G. Falony, Y. Darzi, E. F. Tigchelaar, J. Wang, R. Y. Tito, C. Schiweck, A. Kurilshikov, M. Joossens, C. Wijmenga, S. Claes, L. Van Oudenhove, A. Zhernakova, S. Vieira-Silva, J. Raes, The neuroactive potential of the human gut microbiota in quality of life and depression. *Nat Microbiol.* **4**, 623–632 (2019).

105. E. Rinninella, P. Raoul, M. Cintoni, F. Franceschi, G. A. D. Miggiano, A. Gasbarrini, M. C. Mele, What is the Healthy Gut Microbiota Composition? A Changing Ecosystem across Age, Environment, Diet, and Diseases. *Microorganisms*. **7** (2019), doi:10.3390/microorganisms7010014.

106. D. N. O'Dwyer, R. P. Dickson, B. B. Moore, The Lung Microbiome, Immunity, and the Pathogenesis of Chronic Lung Disease. *J. Immunol.* **196**, 4839–4847 (2016).

107. C. D. Link, Is There a Brain Microbiome? *Neurosci Insights*. **16**, 26331055211018709 (2021).

108. M. C. Collado, S. Rautava, J. Aakko, E. Isolauri, S. Salminen, Human gut colonisation may be initiated in utero by distinct microbial communities in the placenta and amniotic fluid. *Sci. Rep.* **6**, 23129 (2016).

109. P. Ferretti, E. Pasolli, A. Tett, F. Asnicar, V. Gorfer, S. Fedi, F. Armanini, D. T. Truong, S. Manara, M. Zolfo, F. Beghini, R. Bertorelli, V. De Sanctis, I. Bariletti, R. Canto, R. Clementi, M. Cologna, T. Crifò, G. Cusumano, S. Gottardi, C. Innamorati, C. Masè, D. Postai, D. Savoi, S. Duranti, G. A. Lugli, L. Mancabelli, F. Turroni, C. Ferrario, C. Milani, M. Mangifesta, R. Anzalone, A. Viappiani, M. Yassour, H. Vlamakis, R. Xavier, C. M. Collado, O. Koren, S. Tateo, M. Soffiati, A. Pedrotti, M.

Ventura, C. Huttenhower, P. Bork, N. Segata, Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut Microbiome. *Cell Host Microbe.* **24**, 133–145.e5 (2018).

110. G. Berg, D. Rybakova, D. Fischer, T. Cernava, M.-C. C. Vergès, T. Charles, X. Chen, L. Cocolin, K. Eversole, G. H. Corral, M. Kazou, L. Kinkel, L. Lange, N. Lima, A. Loy, J. A. Macklin, E. Maguin, T. Mauchline, R. McClure, B. Mitter, M. Ryan, I. Sarand, H. Smidt, B. Schelkle, H. Roume, G. S. Kiran, J. Selvin, R. S. C. de Souza, L. van Overbeek, B. K. Singh, M. Wagner, A. Walsh, A. Sessitsch, M. Schloter, Microbiome definition re-visited: old concepts and new challenges. *Microbiome.* **8**, 103 (2020).

111. P. Hugenholtz, M. Chuvochina, A. Oren, D. H. Parks, R. M. Soo, Prokaryotic taxonomy and nomenclature in the age of big sequence data. *ISME J.* **15**, 1879–1892 (2021).

112. K. H. Schleifer, Classification of Bacteria and Archaea: past, present and future. *Syst. Appl. Microbiol.* **32**, 533–542 (2009).

113. M. Groussin, M. Poyet, A. Sistiaga, S. M. Kearney, K. Moniz, M. Noel, J. Hooker, S. M. Gibbons, L. Segurel, A. Froment, R. S. Mohamed, A. Fezeu, V. A. Juimo, S. Lafosse, F. E. Tabe, C. Girard, D. Iqaluk, L. T. T. Nguyen, B. J. Shapiro, J. Lehtimäki, L. Ruokolainen, P. P. Kettunen, T. Vatanen, S. Sigwazi, A. Mabulla, M. Domínguez-Rodrigo, Y. A. Nartey, A. Agyei-Nkansah, A. Duah, Y. A. Awuku, K. A. Valles, S. O. Asibey, M. Y. Afihene, L. R. Roberts, A. Plymoth, C. A. Onyekwere, R. E. Summons, R. J. Xavier, E. J. Alm, Elevated rates of horizontal gene transfer in the industrialized human microbiome. *Cell.* **184**, 2053–2067.e18 (2021).

114. A. Oren, G. M. Garrity, Valid publication of the names of forty-two phyla of prokaryotes. *Int. J. Syst. Evol. Microbiol.* **71** (2021), doi:10.1099/ijsem.0.005056.

115. R. Ranjan, A. Rani, A. Metwally, H. S. McGee, D. L. Perkins, Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochem. Biophys. Res. Commun.* **469**, 967–977 (2016).

116. B. A. Clemmons, B. H. Voy, P. R. Myer, Altering the Gut Microbiome of Cattle: Considerations of Host-Microbiome Interactions for Persistent Microbiome Manipulation. *Microb. Ecol.* **77**, 523–536 (2019).

117. A. J. Kessler, Y.-J. Chen, D. W. Waite, T. Hutchinson, S. Koh, M. E. Popa, J. Beardall, P. Hugenholtz, P. L. M. Cook, C. Greening, Bacterial fermentation and respiration processes are uncoupled in anoxic permeable sediments. *Nat Microbiol.* **4**, 1014–1023 (2019).

118. R. Paul Ross, S. Morgan, C. Hill, Preservation and fermentation: past, present and future. *Int. J. Food Microbiol.* **79**, 3–16 (2002).

119. K. Oliphant, E. Allen-Vercoe, Macronutrient metabolism by the human gut microbiome: major fermentation by-products and their impact on host health. *Microbiome.* **7**, 91 (2019).

120. D. J. Morrison, T. Preston, Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. *Gut Microbes.* **7**, 189–200 (2016).

121. R. E. Ley, D. A. Peterson, J. I. Gordon, Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell.* **124**, 837–848 (2006).

122. F. S. Chapin, L. R. Walker, C. L. Fastie, L. C. Sharman, Mechanisms of primary succession

following deglaciation at glacier bay, Alaska. *Ecol. Monogr.* **64**, 149–175 (1994).

123. H. Kozakova, M. Schwarzer, L. Tuckova, D. Srutkova, E. Czarnowska, I. Rosiak, T. Hudcovic, I. Schabussova, P. Hermanova, Z. Zakostelska, T. Aleksandrzak-Piekarczyk, A. Koryszewska-Baginska, H. Tlaskalova-Hogenova, B. Cukrowska, Colonization of germ-free mice with a mixture of three lactobacillus strains enhances the integrity of gut mucosa and ameliorates allergic sensitization. *Cell. Mol. Immunol.* **13**, 251–262 (2016).

124. M. Krishna, S. Gupta, M. Delgado-Baquerizo, E. Morriën, S. C. Garkoti, R. Chaturvedi, S. Ahmad, Successional trajectory of bacterial communities in soil are shaped by plant-driven changes during secondary succession. *Sci. Rep.* **10**, 9864 (2020).

125. G. L. Peterfreund, L. E. Vandivier, R. Sinha, A. J. Marozsan, W. C. Olson, J. Zhu, F. D. Bushman, Succession in the gut microbiome following antibiotic and antibody therapies for Clostridium difficile. *PLoS One.* **7**, e46966 (2012).

126. H. S. Horn, The ecology of secondary succession. *Annu. Rev. Ecol. Syst.* **5**, 25–37 (1974).

127. R. Margalef, "TEMPORAL SUCCESSION AND SPATIAL HETEROGENEITY IN PHYTOPLANKTON" in *Perspectives in Marine Biology* (University of California Press, 1958), pp. 323–350.

128. F. E. Clements, Nature and structure of the climax. *J. Ecol.* **24**, 252 (1936).

129. J. Hu, D. R. Amor, M. Barbier, G. Bunin, J. Gore, Emergent phases of ecological diversity and dynamics mapped in microcosms. *Science.* **378**, 85–89 (2022).

130. M. Arumugam, J. Raes, E. Pelletier, D. Le Paslier, T. Yamada, D. R. Mende, G. R. Fernandes, J. Tap, T. Bruls, J.-M. Batto, M. Bertalan, N. Borruel, F. Casellas, L. Fernandez, L. Gautier, T. Hansen, M. Hattori, T. Hayashi, M. Kleerebezem, K. Kurokawa, M. Leclerc, F. Levenez, C. Manichanh, H. B. Nielsen, T. Nielsen, N. Pons, J. Poulain, J. Qin, T. Sicheritz-Ponten, S. Tims, D. Torrents, E. Ugarte, E. G. Zoetendal, J. Wang, F. Guarner, O. Pedersen, W. M. de Vos, S. Brunak, J. Doré, MetaHIT Consortium, M. Antolín, F. Artiguenave, H. M. Blottiere, M. Almeida, C. Brechot, C. Cara, C. Chervaux, A. Cultrone, C. Delorme, G. Denariaz, R. Dervyn, K. U. Foerstner, C. Friss, M. van de Guchte, E. Guedon, F. Haimet, W. Huber, J. van Hylckama-Vlieg, A. Jamet, C. Juste, G. Kaci, J. Knol, O. Lakhdari, S. Layec, K. Le Roux, E. Maguin, A. Mérieux, R. Melo Minardi, C. M'rini, J. Muller, R. Oozeer, J. Parkhill, P. Renault, M. Rescigno, N. Sanchez, S. Sunagawa, A. Torrejon, K. Turner, G. Vandemeulebrouck, E. Varela, Y. Winogradsky, G. Zeller, J. Weissenbach, S. D. Ehrlich, P. Bork, Enterotypes of the human gut microbiome. *Nature.* **473**, 174–180 (2011).

131. S. Cobo-López, V. K. Gupta, J. Sung, R. Guimerà, M. Sales-Pardo, Stochastic block models reveal a robust nested pattern in healthy human gut microbiomes. *PNAS Nexus.* **1**, gac055 (2022).

132. R. H. Whittaker, Evolution and measurement of species diversity. *Taxon.* **21**, 213–251 (1972).

133. L. Wittebolle, M. Marzorati, L. Clement, A. Balloi, D. Daffonchio, K. Heylen, P. De Vos, W. Verstraete, N. Boon, Initial community evenness favours functionality under selective stress. *Nature.* **458**, 623–626 (2009).

134. F. Finotello, E. Mastrorilli, B. Di Camillo, Measuring the diversity of the human microbiota with targeted next-generation sequencing. *Brief. Bioinform.* **19**, 679–692 (2018).

135. E. Rackaityte, J. Halkias, E. M. Fukui, V. F. Mendoza, C. Hayzelden, E. D. Crawford, K. E.

Fujimura, T. D. Burt, S. V. Lynch, Viable bacterial colonization is highly limited in the human intestine in utero. *Nat. Med.* **26**, 599–607 (2020).

136. M. C. Collado, N. Segata, Initial exploration of in utero microbial colonization. *Nat. Med.* **26** (2020), pp. 469–470.

137. J. M. Rodríguez, K. Murphy, C. Stanton, R. P. Ross, O. I. Kober, N. Juge, E. Avershina, K. Rudi, A. Narbad, M. C. Jenmalm, J. R. Marchesi, M. C. Collado, The composition of the gut microbiota throughout life, with an emphasis on early life. *Microb. Ecol. Health Dis.* **26**, 26050 (2015).

138. R. I. Mackie, A. Sghir, H. R. Gaskins, Developmental microbial ecology of the neonatal gastrointestinal tract. *Am. J. Clin. Nutr.* **69**, 1035S–1045S (1999).

139. C. Martino, A. H. Dilmore, Z. M. Burcham, J. L. Metcalf, D. Jeste, R. Knight, Microbiota succession throughout life from the cradle to the grave. *Nat. Rev. Microbiol.* **20**, 707–720 (2022).

140. M. G. Dominguez-Bello, E. K. Costello, M. Contreras, M. Magris, G. Hidalgo, N. Fierer, R. Knight, Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 11971–11975 (2010).

141. M. G. Dominguez-Bello, K. M. De Jesus-Laboy, N. Shen, L. M. Cox, A. Amir, A. Gonzalez, N. A. Bokulich, S. J. Song, M. Hoashi, J. I. Rivera-Vinas, K. Mendez, R. Knight, J. C. Clemente, Partial restoration of the microbiota of cesarean-born infants via vaginal microbial transfer. *Nat. Med.* **22**, 250–253 (2016).

142. A. Nogacka, N. Salazar, M. Suárez, C. Milani, S. Arboleya, G. Solís, N. Fernández, L. Alaez, A. M. Hernández-Barranco, C. G. de Los Reyes-Gavilán, M. Ventura, M. Gueimonde, Impact of intrapartum antimicrobial prophylaxis upon the intestinal microbiota and the prevalence of antibiotic resistance genes in vaginally delivered full-term neonates. *Microbiome.* **5**, 93 (2017).

143. S. Ainonen, M. V. Tejesvi, M. R. Mahmud, N. Paalanne, T. Pokka, W. Li, K. E. Nelson, J. Salo, M. Renko, P. Vänni, A. M. Pirttilä, T. Tapiainen, Antibiotics at birth and later antibiotic courses: effects on gut microbiota. *Pediatr. Res.* **91**, 154–162 (2022).

144. M. Reyman, M. A. van Houten, R. L. Watson, M. L. J. N. Chu, K. Arp, W. J. de Waal, I. Schiering, F. B. Plötz, R. J. L. Willems, W. van Schaik, E. A. M. Sanders, D. Bogaert, Effects of early-life antibiotics on the developing infant gut microbiome and resistome: a randomized trial. *Nat. Commun.* **13**, 893 (2022).

145. A. Parra-Llorca, M. Gormaz, C. Alcántara, M. Cernada, A. Nuñez-Ramiro, M. Vento, M. C. Collado, Preterm Gut Microbiome Depending on Feeding Type: Significance of Donor Human Milk. *Front. Microbiol.* **9**, 1376 (2018).

146. S. M. Parigi, M. Eldh, P. Larssen, S. Gabrielsson, E. J. Villablanca, Breast Milk and Solid Food Shaping Intestinal Immunity. *Front. Immunol.* **6**, 415 (2015).

147. M. K. McGuire, M. A. McGuire, Got bacteria? The astounding, yet not-so-surprising, microbiome of human milk. *Curr. Opin. Biotechnol.* **44**, 63–68 (2017).

148. Q. S. Damaceno, J. P. Souza, J. R. Nicoli, R. L. Paula, G. B. Assis, H. C. Figueiredo, V. Azevedo, F. S. Martins, Evaluation of Potential Probiotics Isolated from Human Milk and Colostrum. *Probiotics Antimicrob. Proteins*. **9**, 371–379 (2017).

149. J. E. Bisanz, M. K. Enos, G. PrayGod, S. Seney, J. M. Macklaim, S. Chilton, D. Willner, R. Knight, C. Fusch, G. Fusch, G. B. Gloor, J. P. Burton, G. Reid, Microbiota at Multiple Body Sites during Pregnancy in a Rural Tanzanian Population and Effects of Moringa-Supplemented Probiotic Yogurt. *Appl. Environ. Microbiol.* **81**, 4965–4975 (2015).

150. S. Moossavi, S. Sepehri, B. Robertson, L. Bode, S. Goruk, C. J. Field, L. M. Lix, R. J. de Souza, A. B. Becker, P. J. Mandhane, S. E. Turvey, P. Subbarao, T. J. Moraes, D. L. Lefebvre, M. R. Sears, E. Khafipour, M. B. Azad, Composition and Variation of the Human Milk Microbiota Are Influenced by Maternal and Early-Life Factors. *Cell Host Microbe.* **25**, 324–335.e4 (2019).

151. A. Marcobal, M. Barboza, J. W. Froehlich, D. E. Block, J. B. German, C. B. Lebrilla, D. A. Mills, Consumption of human milk oligosaccharides by gut-related microbes. *J. Agric. Food Chem.* **58**, 5334–5340 (2010).

152. G. M. Ruiz-Palacios, L. E. Cervantes, P. Ramos, B. Chavez-Munguia, D. S. Newburg, Campylobacter jejuni binds intestinal H(O) antigen (Fuc alpha 1, 2Gal beta 1, 4GlcNAc), and fucosyloligosaccharides of human milk inhibit its binding and infection. *J. Biol. Chem.* **278**, 14112–14120 (2003).

153. L. Bode, Human milk oligosaccharides: every baby needs a sugar mama. *Glycobiology.* **22**, 1147–1162 (2012).

154. M. B. Azad, T. Konya, R. R. Persaud, D. S. Guttman, R. S. Chari, C. J. Field, M. R. Sears, P. J. Mandhane, S. E. Turvey, P. Subbarao, A. B. Becker, J. A. Scott, A. L. Kozyrskyj, CHILD Study Investigators, Impact of maternal intrapartum antibiotics, method of birth and breastfeeding on gut microbiota during the first year of life: a prospective cohort study. *BJOG.* **123**, 983–993 (2016).

155. M. Fallani, S. Amarri, A. Uusijarvi, R. Adam, S. Khanna, M. Aguilera, A. Gil, J. M. Vieites, E. Norin, D. Young, J. A. Scott, J. Doré, C. A. Edwards, The Infabio Team, Determinants of the human infant intestinal microbiota after the introduction of first complementary foods in infant samples from five European centres. *Microbiology.* **157**, 1385–1392 (2011).

156. M. Derrien, A.-S. Alvarez, W. M. de Vos, The Gut Microbiota in the First Decade of Life. *Trends Microbiol.* **27**, 997–1010 (2019).

157. T. Yatsunenko, F. E. Rey, M. J. Manary, I. Trehan, M. G. Dominguez-Bello, M. Contreras, M. Magris, G. Hidalgo, R. N. Baldassano, A. P. Anokhin, A. C. Heath, B. Warner, J. Reeder, J. Kuczynski, J. G. Caporaso, C. A. Lozupone, C. Lauber, J. C. Clemente, D. Knights, R. Knight, J. I. Gordon, Human gut microbiome viewed across age and geography. *Nature.* **486**, 222–227 (2012).

158. J. E. Koenig, A. Spor, N. Scalfone, A. D. Fricker, J. Stombaugh, R. Knight, L. T. Angenent, R. E. Ley, Succession of microbial consortia in the developing infant gut microbiome. *Proc. Natl. Acad. Sci. U. S. A.* **108 Suppl 1**, 4578–4585 (2011).

159. J. C. Clemente, E. C. Pehrsson, M. J. Blaser, K. Sandhu, Z. Gao, B. Wang, M. Magris, G. Hidalgo, M. Contreras, Ó. Noya-Alarcón, O. Lander, J. McDonald, M. Cox, J. Walter, P. L. Oh, J. F. Ruiz, S. Rodriguez, N. Shen, S. J. Song, J. Metcalf, R. Knight, G. Dantas, M. G. Dominguez-Bello, The microbiome of uncontacted Amerindians. *Sci Adv.* **1** (2015), doi:10.1126/sciadv.1500183.

160. S. L. Schnorr, M. Candela, S. Rampelli, M. Centanni, C. Consolandi, G. Basaglia, S. Turroni, E. Biagi, C. Peano, M. Severgnini, J. Fiori, R. Gotti, G. De Bellis, D. Luiselli, P. Brigidi, A. Mabulla, F. Marlowe, A. G. Henry, A. N. Crittenden, Gut microbiome of the Hadza hunter-gatherers. *Nat.*

Commun. **5**, 3654 (2014).

161. D. Mariat, O. Firmesse, F. Levenez, V. Guimarăes, H. Sokol, J. Doré, G. Corthier, J.-P. Furet, The Firmicutes/Bacteroidetes ratio of the human microbiota changes with age. *BMC Microbiol.* **9**, 123 (2009).

162. M. Vacca, B. Raspini, F. M. Calabrese, D. Porri, R. De Giuseppe, M. Chieppa, M. Liso, R. M. Cerbo, E. Civardi, F. Garofoli, H. Cena, M. De Angelis, The establishment of the gut microbiota in 1-year-aged infants: from birth to family food. *Eur. J. Nutr.* **61**, 2517–2530 (2022).

163. C.-M. Homann, C. A. J. Rossel, S. Dizzell, L. Bervoets, J. Simioni, J. Li, E. Gunn, M. G. Surette, R. J. de Souza, M. Mommers, E. K. Hutton, K. M. Morrison, J. Penders, N. van Best, J. C. Stearns, Infants' First Solid Foods: Impact on Gut Microbiota Development in Two Intercontinental Cohorts. *Nutrients*. **13** (2021), doi:10.3390/nu13082639.

164. N. Ottman, H. Smidt, W. M. de Vos, C. Belzer, The function of our microbiota: who is out there and what do they do? *Front. Cell. Infect. Microbiol.* **2**, 104 (2012).

165. F. Bishehsari, R. M. Voigt, A. Keshavarzian, Circadian rhythms and the gut microbiota: from the metabolic syndrome to cancer. *Nat. Rev. Endocrinol.* **16**, 731–739 (2020).

166. I. C. Mason, J. Qian, G. K. Adler, F. A. J. L. Scheer, Impact of circadian disruption on glucose metabolism: implications for type 2 diabetes. *Diabetologia*. **63**, 462–472 (2020).

167. S. S. Thosar, M. P. Butler, S. A. Shea, Role of the circadian system in cardiovascular disease. *J. Clin. Invest.* **128**, 2157–2167 (2018).

168. I. N. Karatsoreos, Links between Circadian Rhythms and Psychiatric Disease. *Front. Behav. Neurosci.* **8**, 162 (2014).

169. G. Sulli, M. T. Y. Lam, S. Panda, Interplay between Circadian Clock and Cancer: New Frontiers for Cancer Treatment. *Trends Cancer Res.* **5**, 475–494 (2019).

170. B. D. Weger, C. Gobet, J. Yeung, E. Martin, S. Jimenez, B. Betrisey, F. Foata, B. Berger, A. Balvay, A. Foussier, A. Charpagne, B. Boizet-Bonhoure, C. J. Chou, F. Naef, F. Gachon, The Mouse Microbiome Is Required for Sex-Specific Diurnal Rhythms of Gene Expression and Metabolism. *Cell Metab.* **29**, 362–382.e8 (2019).

171. S. A. Smits, J. Leach, E. D. Sonnenburg, C. G. Gonzalez, J. S. Lichtman, G. Reid, R. Knight, A. Manjurano, J. Changalucha, J. E. Elias, M. G. Dominguez-Bello, J. L. Sonnenburg, Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania. *Science*. **357**, 802–806 (2017).

172. L. Kešnerová, O. Emery, M. Troilo, J. Liberti, B. Erkosar, P. Engel, Gut microbiota structure differs between honeybees in winter and summer. *ISME J.* **14**, 801–814 (2020).

173. L. A. David, C. F. Maurice, R. N. Carmody, D. B. Gootenberg, J. E. Button, B. E. Wolfe, A. V. Ling, A. S. Devlin, Y. Varma, M. A. Fischbach, S. B. Biddinger, R. J. Dutton, P. J. Turnbaugh, Diet rapidly and reproducibly alters the human gut microbiome. *Nature*. **505**, 559–563 (2014).

174. C. Losasso, E. M. Eckert, E. Mastrorilli, J. Villiger, M. Mancin, I. Patuzzi, A. Di Cesare, V. Cibin, F. Barrucci, J. Pernthaler, G. Corno, A. Ricci, Assessing the influence of vegan, vegetarian and omnivore oriented westernized dietary styles on human gut Microbiota: A cross sectional study. *Front. Microbiol.* **9**, 317 (2018).

175. A. Tomova, I. Bukovsky, E. Rembert, W. Yonas, J. Alwarith, N. D. Barnard, H. Kahleova, The Effects of Vegetarian and Vegan Diets on Gut Microbiota. *Front Nutr.* **6**, 47 (2019).

176. B. B. Matijašić, T. Obermajer, L. Lipoglavšek, I. Grabnar, G. Avguštin, I. Rogelj, Association of dietary type with fecal microbiota in vegetarians and omnivores in Slovenia. *Eur. J. Nutr.* **53**, 1051–1064 (2014).

177. F. De Filippis, N. Pellegrini, L. Vannini, I. B. Jeffery, A. La Storia, L. Laghi, D. I. Serrazanetti, R. Di Cagno, I. Ferrocino, C. Lazzi, S. Turroni, L. Cocolin, P. Brigidi, E. Neviani, M. Gobbetti, P. W. O'Toole, D. Ercolini, High-level adherence to a Mediterranean diet beneficially impacts the gut microbiota and associated metabolome. *Gut.* **65**, 1812–1821 (2016).

178. E. D. Sonnenburg, J. L. Sonnenburg, The ancestral and industrialized gut microbiota and implications for human health. *Nat. Rev. Microbiol.* **17**, 383–390 (2019).

179. M. P. Francino, Antibiotics and the Human Gut Microbiome: Dysbioses and Accumulation of Resistances. *Front. Microbiol.* **6**, 1543 (2015).

180. R. Vivas, A. A. T. Barbosa, S. S. Dolabela, S. Jain, Multidrug-Resistant Bacteria and Alternative Methods to Control Them: An Overview. *Microb. Drug Resist.* **25**, 890–908 (2019).

181. D. Romero-Calle, R. Guimarães Benevides, A. Góes-Neto, C. Billington, Bacteriophages as Alternatives to Antibiotics in Clinical Care. *Antibiotics (Basel)*. **8** (2019), doi:10.3390/antibiotics8030138.

182. J.-H. Shin, Y.-H. Park, M. Sim, S.-A. Kim, H. Joung, D.-M. Shin, Serum level of sex steroid hormone is associated with diversity and profiles of human gut microbiome. *Res. Microbiol.* **170**, 192–201 (2019).

183. H. Colldén, A. Landin, V. Wallenius, E. Elebring, L. Fändriks, M. E. Nilsson, H. Ryberg, M. Poutanen, K. Sjögren, L. Vandenput, C. Ohlsson, The gut microbiota is a major regulator of androgen metabolism in intestinal contents. *Am. J. Physiol. Endocrinol. Metab.* **317**, E1182–E1192 (2019).

184. S. Liu, R. Cao, L. Liu, Y. Lv, X. Qi, Z. Yuan, X. Fan, C. Yu, Q. Guan, Correlation Between Gut Microbiota and Testosterone in Male Patients With Type 2 Diabetes Mellitus. *Front. Endocrinol.* . **13**, 836485 (2022).

185. J. M. Baker, L. Al-Nakkash, M. M. Herbst-Kralovetz, Estrogen-gut microbiome axis: Physiological and clinical implications. *Maturitas*. **103**, 45–53 (2017).

186. P.-H. Wang, Y.-L. Chen, S. T.-S. Wei, K. Wu, T.-H. Lee, T.-Y. Wu, Y.-R. Chiang, Retroconversion of estrogens into androgens by bacteria via a cobalamin-mediated methylation. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 1395–1403 (2020).

187. J. G. M. Markle, D. N. Frank, S. Mortin-Toth, C. E. Robertson, L. M. Feazel, U. Rolle-Kampczyk, M. von Bergen, K. D. McCoy, A. J. Macpherson, J. S. Danska, Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity. *Science*. **339**, 1084–1088 (2013).

188. Y. Qin, A. S. Havulinna, Y. Liu, P. Jousilahti, S. C. Ritchie, A. Tokolyi, J. G. Sanders, L. Valsta, M. Brożyńska, Q. Zhu, A. Tripathi, Y. Vázquez-Baeza, R. Loomba, S. Cheng, M. Jain, T. Niiranen, L. Lahti, R. Knight, V. Salomaa, M. Inouye, G. Méric, Combined effects of host genetics and diet on human gut microbiota and incident disease in a single population cohort. *Nat. Genet.* **54**, 134–142

(2022).

189. B. Darabi, S. Rahmati, M. R. HafeziAhmadi, G. Badfar, M. Azami, The association between caesarean section and childhood asthma: an updated systematic review and meta-analysis. *Allergy Asthma Clin. Immunol.* **15**, 62 (2019).

190. N. T. Mueller, R. Whyatt, L. Hoepner, S. Oberfield, M. G. Dominguez-Bello, E. M. Widen, A. Hassoun, F. Perera, A. Rundle, Prenatal exposure to antibiotics, cesarean section and risk of childhood obesity. *Int. J. Obes.* . **39**, 665–670 (2015).

191. C. R. Cardwell, L. C. Stene, G. Joner, O. Cinek, J. Svensson, M. J. Goldacre, R. C. Parslow, P. Pozzilli, G. Brigis, D. Stoyanov, B. Urbonaite, S. Sipetić, E. Schober, C. Ionescu-Tirgoviste, G. Devoti, C. E. de Beaufort, K. Buschard, C. C. Patterson, Caesarean section is associated with an increased risk of childhood-onset type 1 diabetes mellitus: a meta-analysis of observational studies. *Diabetologia*. **51**, 726–735 (2008).

192. M. Trøseid, G. Ø. Andersen, K. Broch, J. R. Hov, The gut microbiome in coronary artery disease and heart failure: Current knowledge and future directions. *EBioMedicine*. **52**, 102649 (2020).

193. V. L. Nikolova, M. R. B. Smith, L. J. Hall, A. J. Cleare, J. M. Stone, A. H. Young, Perturbations in Gut Microbiota Composition in Psychiatric Disorders: A Review and Meta-analysis. *JAMA Psychiatry*. **78**, 1343–1354 (2021).

194. C. Jiang, G. Li, P. Huang, Z. Liu, B. Zhao, The Gut Microbiota and Alzheimer's Disease. *J. Alzheimers. Dis.* **58**, 1–15 (2017).

195. J. A. Gilbert, R. Krajmalnik-Brown, D. L. Porazinska, S. J. Weiss, R. Knight, Toward effective probiotics for autism and other neurodevelopmental disorders. *Cell*. **155** (2013), pp. 1446–1448.

196. S. H. Wong, J. Yu, Gut microbiota in colorectal cancer: mechanisms of action and clinical applications. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 690–704 (2019).

197. Z. Eslami-S, K. Majidzadeh-A, S. Halvaei, F. Babapirali, R. Esmaeili, Microbiome and Breast Cancer: New Role for an Ancient Population. *Front. Oncol.* **10**, 120 (2020).

198. F. De Luca, Y. Shoenfeld, The microbiome in autoimmune diseases. *Clin. Exp. Immunol.* **195**, 74–85 (2019).

199. R. A. Koeth, Z. Wang, B. S. Levison, J. A. Buffa, E. Org, B. T. Sheehy, E. B. Britt, X. Fu, Y. Wu, L. Li, J. D. Smith, J. A. DiDonato, J. Chen, H. Li, G. D. Wu, J. D. Lewis, M. Warrier, J. M. Brown, R. M. Krauss, W. H. W. Tang, F. D. Bushman, A. J. Lusis, S. L. Hazen, Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat. Med.* **19**, 576–585 (2013).

200. D. Houghton, C. J. Stewart, C. Stamp, A. Nelson, N. J. Aj Ami, J. F. Petrosino, A. Wipat, M. I. Trenell, D. M. Turnbull, L. C. Greaves, Impact of Age-Related Mitochondrial Dysfunction and Exercise on Intestinal Microbiota Composition. *J. Gerontol. A Biol. Sci. Med. Sci.* **73**, 571–578 (2018).

201. A. Clark, N. Mach, The Crosstalk between the Gut Microbiota and Mitochondria during Exercise. *Front. Physiol.* **8**, 319 (2017).

202. D. A. Hughes, R. Bacigalupe, J. Wang, M. C. Rühlemann, R. Y. Tito, G. Falony, M. Joossens, S.

Vieira-Silva, L. Henckaerts, L. Rymenans, C. Verspecht, S. Ring, A. Franke, K. H. Wade, N. J. Timpson, J. Raes, Genome-wide associations of human gut microbiome variation and implications for causal inference analyses. *Nat Microbiol.* **5**, 1079–1087 (2020).

203. R. Estruch, E. Ros, J. Salas-Salvadó, M.-I. Covas, D. Corella, F. Arós, E. Gómez-Gracia, V. Ruiz-Gutiérrez, M. Fiol, J. Lapetra, R. M. Lamuela-Raventos, L. Serra-Majem, X. Pintó, J. Basora, M. A. Muñoz, J. V. Sorlí, J. A. Martínez, M. A. Martínez-González, PREDIMED Study Investigators, Primary prevention of cardiovascular disease with a Mediterranean diet. *N. Engl. J. Med.* **368**, 1279–1290 (2013).

204. E. H. Martínez-Lapiscina, P. Clavero, E. Toledo, R. Estruch, J. Salas-Salvadó, B. San Julián, A. Sanchez-Tainta, E. Ros, C. Valls-Pedret, M. Á. Martinez-Gonzalez, Mediterranean diet improves cognition: the PREDIMED-NAVARRA randomised trial. *J. Neurol. Neurosurg. Psychiatry*. **84**, 1318–1325 (2013).

205. M. T. Mitjavila, M. Fandos, J. Salas-Salvadó, M.-I. Covas, S. Borrego, R. Estruch, R. Lamuela-Raventós, D. Corella, M. Á. Martínez-Gonzalez, J. M. Sánchez, M. Bulló, M. Fitó, C. Tormos, C. Cerdá, R. Casillas, J. J. Moreno, A. Iradi, C. Zaragoza, J. Chaves, G. T. Sáez, The Mediterranean diet improves the systemic lipid and DNA oxidative damage in metabolic syndrome individuals. A randomized, controlled, trial. *Clin. Nutr.* **32**, 172–178 (2013).

206. D. Zeevi, T. Korem, N. Zmora, D. Israeli, D. Rothschild, A. Weinberger, O. Ben-Yacov, D. Lador, T. Avnit-Sagi, M. Lotan-Pompan, J. Suez, J. A. Mahdi, E. Matot, G. Malka, N. Kosower, M. Rein, G. Zilberman-Schapira, L. Dohnalová, M. Pevsner-Fischer, R. Bikovsky, Z. Halpern, E. Elinav, E. Segal, Personalized Nutrition by Prediction of Glycemic Responses. *Cell.* **163**, 1079–1094 (2015).

207. H. C. Wastyk, G. K. Fragiadakis, D. Perelman, D. Dahan, B. D. Merrill, F. B. Yu, M. Topf, C. G. Gonzalez, W. Van Treuren, S. Han, J. L. Robinson, J. E. Elias, E. D. Sonnenburg, C. D. Gardner, J. L. Sonnenburg, Gut-microbiota-targeted diets modulate human immune status. *Cell.* **184**, 4137–4153.e14 (2021).

208. S. J. Oak, R. Jha, The effects of probiotics in lactose intolerance: A systematic review. *Crit. Rev. Food Sci. Nutr.* **59**, 1675–1683 (2019).

209. A. Guarino, S. Guandalini, A. Lo Vecchio, Probiotics for Prevention and Treatment of Diarrhea. *J. Clin. Gastroenterol.* **49 Suppl 1**, S37–45 (2015).

210. I. Bjarnason, G. Sission, B. 'hussaine Hayee, A randomised, double-blind, placebo-controlled trial of a multi-strain probiotic in patients with asymptomatic ulcerative colitis and Crohn's disease. *Inflammopharmacology*. **27**, 465–473 (2019).

211. K. Lindfors, T. Blomqvist, K. Juuti-Uusitalo, S. Stenman, J. Venäläinen, M. Mäki, K. Kaukinen, Live probiotic Bifidobacterium lactis bacteria inhibit the toxic effects induced by wheat gliadin in epithelial cell culture. *Clin. Exp. Immunol.* **152**, 552–558 (2008).

212. M. X. Maldonado-Gómez, I. Martínez, F. Bottacini, A. O'Callaghan, M. Ventura, D. van Sinderen, B. Hillmann, P. Vangay, D. Knights, R. W. Hutkins, J. Walter, Stable Engraftment of Bifidobacterium longum AH1206 in the Human Gut Depends on Individualized Features of the Resident Microbiome. *Cell Host Microbe.* **20**, 515–526 (2016).

213. J. Walter, M. X. Maldonado-Gómez, I. Martínez, To engraft or not to engraft: an ecological framework for gut microbiome modulation with live microbes. *Curr. Opin. Biotechnol.* **49**, 129–139

(2018).

214. H. D. Holscher, Dietary fiber and prebiotics and the gastrointestinal microbiota. *Gut Microbes*. **8**, 172–184 (2017).

215. J.-W. Huaman, M. Mego, C. Manichanh, N. Cañellas, D. Cañueto, H. Segurola, M. Jansana, C. Malagelada, A. Accarino, J. Vulevic, G. Tzortzis, G. Gibson, E. Saperas, F. Guarner, F. Azpiroz, Effects of Prebiotics vs a Diet Low in FODMAPs in Patients With Functional Gut Disorders. *Gastroenterology*. **155**, 1004–1007 (2018).

216. B. Wilson, M. Rossi, E. Dimidi, K. Whelan, Prebiotics in irritable bowel syndrome and other functional bowel disorders in adults: a systematic review and meta-analysis of randomized controlled trials. *Am. J. Clin. Nutr.* **109**, 1098–1111 (2019).

217. S. Salminen, M. C. Collado, A. Endo, C. Hill, S. Lebeer, E. M. M. Quigley, M. E. Sanders, R. Shamir, J. R. Swann, H. Szajewska, G. Vinderola, The International Scientific Association of Probiotics and Prebiotics (ISAPP) consensus statement on the definition and scope of postbiotics. *Nat. Rev. Gastroenterol. Hepatol.* **18**, 649–667 (2021).

218. K. S. Swanson, G. R. Gibson, R. Hutkins, R. A. Reimer, G. Reid, K. Verbeke, K. P. Scott, H. D. Holscher, M. B. Azad, N. M. Delzenne, M. E. Sanders, The International Scientific Association for Probiotics and Prebiotics (ISAPP) consensus statement on the definition and scope of synbiotics. *Nat. Rev. Gastroenterol. Hepatol.* **17**, 687–701 (2020).

219. C. A. M. Wegh, S. Y. Geerlings, J. Knol, G. Roeselers, C. Belzer, Postbiotics and Their Potential Applications in Early Life Nutrition and Beyond. *Int. J. Mol. Sci.* **20** (2019), doi:10.3390/ijms20194673.

220. S. Khanna, M. Bishnoi, K. K. Kondepudi, G. Shukla, Synbiotic (Lactiplantibacillus pentosus GSSK2 and isomalto-oligosaccharides) supplementation modulates pathophysiology and gut dysbiosis in experimental metabolic syndrome. *Sci. Rep.* **11**, 21397 (2021).

221. D. Kao, B. Roach, M. Silva, P. Beck, K. Rioux, G. G. Kaplan, H.-J. Chang, S. Coward, K. J. Goodman, H. Xu, K. Madsen, A. Mason, G. K.-S. Wong, J. Jovel, J. Patterson, T. Louie, Effect of Oral Capsule- vs Colonoscopy-Delivered Fecal Microbiota Transplantation on Recurrent Clostridium difficile Infection: A Randomized Clinical Trial. *JAMA*. **318**, 1985–1993 (2017).

222. E. W. Yu, L. Gao, P. Stastka, M. C. Cheney, J. Mahabamunuge, M. Torres Soto, C. B. Ford, J. A. Bryant, M. R. Henn, E. L. Hohmann, Fecal microbiota transplantation for the improvement of metabolism in obesity: The FMT-TRIM double-blind placebo-controlled pilot trial. *PLoS Med.* **17**, e1003051 (2020).

223. J. P. K. Doll, J. F. Vázquez-Castellanos, A.-C. Schaub, N. Schweinfurth, C. Kettelhack, E. Schneider, G. Yamanbaeva, L. Mählmann, S. Brand, C. Beglinger, S. Borgwardt, J. Raes, A. Schmidt, U. E. Lang, Fecal Microbiota Transplantation (FMT) as an Adjunctive Therapy for Depression-Case Report. *Front. Psychiatry*. **13**, 815422 (2022).

224. S. R. Fehily, C. Basnayake, E. K. Wright, M. A. Kamm, Fecal microbiota transplantation therapy in Crohn's disease: Systematic review. *J. Gastroenterol. Hepatol.* **36**, 2672–2686 (2021).

225. N. Li, H. Chen, Y. Cheng, F. Xu, G. Ruan, S. Ying, W. Tang, L. Chen, M. Chen, L. Lv, Y. Ping, D. Chen, Y. Wei, Fecal Microbiota Transplantation Relieves Gastrointestinal and Autism Symptoms by Improving the Gut Microbiota in an Open-Label Study. *Front. Cell. Infect. Microbiol.* **11**, 759435

(2021).

226. E. N. Baruch, I. Youngster, G. Ben-Betzalel, R. Ortenberg, A. Lahat, L. Katz, K. Adler, D. Dick-Necula, S. Raskin, N. Bloch, D. Rotin, L. Anafi, C. Avivi, J. Melnichenko, Y. Steinberg-Silman, R. Mamtani, H. Harati, N. Asher, R. Shapira-Frommer, T. Brosh-Nissimov, Y. Eshet, S. Ben-Simon, O. Ziv, M. A. W. Khan, M. Amit, N. J. Ajami, I. Barshack, J. Schachter, J. A. Wargo, O. Koren, G. Markel, B. Boursi, Fecal microbiota transplant promotes response in immunotherapy-refractory melanoma patients. *Science.* **371**, 602–609 (2021).

227. Y. Fan, O. Pedersen, Gut microbiota in human metabolic health and disease. *Nat. Rev. Microbiol.* **19**, 55–71 (2021).

228. A. Albillos, A. de Gottardi, M. Rescigno, The gut-liver axis in liver disease: Pathophysiological basis for therapy. *J. Hepatol.* **72**, 558–577 (2020).

229. Y. Belkaid, T. W. Hand, Role of the microbiota in immunity and inflammation. *Cell.* **157**, 121–141 (2014).

230. P. J. Turnbaugh, R. E. Ley, M. Hamady, C. M. Fraser-Liggett, R. Knight, J. I. Gordon, The human microbiome project. *Nature.* **449**, 804–810 (2007).

231. G. Clarke, R. M. Stilling, P. J. Kennedy, C. Stanton, J. F. Cryan, T. G. Dinan, Minireview: Gut microbiota: the neglected endocrine organ. *Mol. Endocrinol.* **28**, 1221–1238 (2014).

232. D. Erny, N. Dokalis, C. Mezö, A. Castoldi, O. Mossad, O. Staszewski, M. Frosch, M. Villa, V. Fuchs, A. Mayer, J. Neuber, J. Sosat, S. Tholen, O. Schilling, A. Vlachos, T. Blank, M. Gomez de Agüero, A. J. Macpherson, E. J. Pearce, M. Prinz, Microbiota-derived acetate enables the metabolic fitness of the brain innate immune system during health and disease. *Cell Metab.* **33**, 2260–2276.e7 (2021).

233. J. M. Ridlon, D. J. Kang, P. B. Hylemon, J. S. Bajaj, Bile acids and the gut microbiome. *Curr. Opin. Gastroenterol.* **30**, 332–338 (2014).

234. T. C. Fung, The microbiota-immune axis as a central mediator of gut-brain communication. *Neurobiol. Dis.* **136**, 104714 (2020).

235. A. Jacobson, D. Yang, M. Vella, I. M. Chiu, The intestinal neuro-immune axis: crosstalk between neurons, immune cells, and microbes. *Mucosal Immunol.* **14**, 555–565 (2021).

236. E. Holmes, J. V. Li, T. Athanasiou, H. Ashrafian, J. K. Nicholson, Understanding the role of gut microbiome-host metabolic signal disruption in health and disease. *Trends Microbiol.* **19**, 349–359 (2011).

237. C. H. Waddington, Canalization of development and the inheritance of acquired characters. *Nature.* **150**, 563–565 (1942).

238. J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick,

V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y. H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J. F. Abril, R. Guigó, M. J. Campbell, K. V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y. H. Chiang, M. Coyne, C. Dahlke, A. Deslattes Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh, X. Zhu, The sequence of the human genome. *Science*. **291**, 1304–1351 (2001).

239. R. A. Harris, T. Wang, C. Coarfa, R. P. Nagarajan, C. Hong, S. L. Downey, B. E. Johnson, S. D. Fouse, A. Delaney, Y. Zhao, A. Olshen, T. Ballinger, X. Zhou, K. J. Forsberg, J. Gu, L. Echipare, H. O'Geen, R. Lister, M. Pelizzola, Y. Xi, C. B. Epstein, B. E. Bernstein, R. D. Hawkins, B. Ren, W.-Y. Chung, H. Gu, C. Bock, A. Gnirke, M. Q. Zhang, D. Haussler, J. R. Ecker, W. Li, P. J. Farnham, R. A. Waterland, A. Meissner, M. A. Marra, M. Hirst, A. Milosavljevic, J. F. Costello, Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nat. Biotechnol.* **28**, 1097–1105 (2010).

240. E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, J. Dekker, Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. **326**, 289–293 (2009).

241. M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, E. Charpentier, A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. **337**, 816–821 (2012).

242. L. Statello, C.-J. Guo, L.-L. Chen, M. Huarte, Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* **22**, 96–118 (2021).

243. M. Lawrence, S. Daujat, R. Schneider, Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends Genet.* **32**, 42–56 (2016).

244. A. Nieborak, R. Schneider, Metabolic intermediates - Cellular messengers talking to chromatin

modifiers. *Mol Metab*. **14**, 39–52 (2018).

245. A. J. Sood, C. Viner, M. M. Hoffman, DNAmod: the DNA modification database. *J. Cheminform.* **11**, 30 (2019).

246. R. Shapiro, R. E. Servis, M. Welcher, Reactions of uracil and cytosine derivatives with sodium bisulfite. *J. Am. Chem. Soc.* **92**, 422–424 (1970).

247. H. Hayatsu, Y. Wataya, K. Kazushige, The addition of sodium bisulfite to uracil and to cytosine. *J. Am. Chem. Soc.* **92**, 724–726 (1970).

248. J. G. Herman, J. R. Graff, S. Myöhänen, B. D. Nelkin, S. B. Baylin, Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 9821–9826 (1996).

249. M. J. Booth, M. R. Branco, G. Ficz, D. Oxley, F. Krueger, W. Reik, S. Balasubramanian, Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science*. **336**, 934–937 (2012).

250. T. H. Bestor, J. R. Edwards, M. Boulard, Notes on the role of dynamic DNA methylation in mammalian development. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 6796–6799 (2015).

251. J.-W. Cho, H. S. Shim, C. Y. Lee, S. Y. Park, M. H. Hong, I. Lee, H. R. Kim, The importance of enhancer methylation for epigenetic regulation of tumorigenesis in squamous lung cancer. *Exp. Mol. Med.* **54**, 12–22 (2022).

252. D. Anastasiadi, A. Esteve-Codina, F. Piferrer, Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. *Epigenetics Chromatin.* **11**, 37 (2018).

253. M. R. Branco, G. Ficz, W. Reik, Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat. Rev. Genet.* **13**, 7–13 (2011).

254. R. J. Klose, A. P. Bird, Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.* **31**, 89–97 (2006).

255. T. B. Johnson, R. D. Coghill, Researches on pyrimidines. C111. The discovery of 5-methyl-cytosine in tuberculinic acid, the nucleic acid of the tubercle bacillus1. *J. Am. Chem. Soc.* **47**, 2838–2844 (1925).

256. A. L. Mattei, N. Bailly, A. Meissner, DNA methylation: a historical perspective. *Trends Genet.* **38**, 676–707 (2022).

257. P. A. Jones, Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13**, 484–492 (2012).

258. R. Lister, M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q.-M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren, J. R. Ecker, Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*. **462**, 315–322 (2009).

259. A. Meissner, T. S. Mikkelsen, H. Gu, M. Wernig, J. Hanna, A. Sivachenko, X. Zhang, B. E. Bernstein, C. Nusbaum, D. B. Jaffe, A. Gnirke, R. Jaenisch, E. S. Lander, Genome-scale DNA

methylation maps of pluripotent and differentiated cells. *Nature*. **454**, 766–770 (2008).

260. N. Bhutani, D. M. Burns, H. M. Blau, DNA demethylation dynamics. *Cell*. **146**, 866–872 (2011).

261. M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*. **324**, 930–935 (2009).

262. C.-X. Song, K. E. Szulwach, Q. Dai, Y. Fu, S.-Q. Mao, L. Lin, C. Street, Y. Li, M. Poidevin, H. Wu, J. Gao, P. Liu, L. Li, G.-L. Xu, P. Jin, C. He, Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell*. **153**, 678–691 (2013).

263. L. Wang, Y. Zhou, L. Xu, R. Xiao, X. Lu, L. Chen, J. Chong, H. Li, C. He, X.-D. Fu, D. Wang, Molecular basis for 5-carboxycytosine recognition by RNA polymerase II elongation complex. *Nature*. **523**, 621–625 (2015).

264. X. Lu, C.-X. Song, K. Szulwach, Z. Wang, P. Weidenbacher, P. Jin, C. He, Chemical modification-assisted bisulfite sequencing (CAB-Seq) for 5-carboxylcytosine detection in DNA. *J. Am. Chem. Soc.* **135**, 9315–9317 (2013).

265. K. Boulias, E. L. Greer, Means, mechanisms and consequences of adenine methylation in DNA. *Nat. Rev. Genet.* **23**, 411–428 (2022).

266. M. Berney, J. F. McGouran, Methods for detection of cytosine and thymine modifications in DNA. *Nat. Rev. Chem.* **2**, 332–348 (2018).

267. A. M. Fleming, J. Zhu, Y. Ding, S. Esders, C. J. Burrows, Oxidative Modification of Guanine in a Potential Z-DNA-Forming Sequence of a Gene Promoter Impacts Gene Expression. *Chem. Res. Toxicol.* **32**, 899–909 (2019).

268. J. L. Workman, R. E. Kingston, Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annu. Rev. Biochem.* **67**, 545–579 (1998).

269. B. Li, M. Carey, J. L. Workman, The role of chromatin during transcription. *Cell*. **128**, 707–719 (2007).

270. T. Jenuwein, C. D. Allis, Translating the histone code. *Science*. **293**, 1074–1080 (2001).

271. A. J. Bannister, T. Kouzarides, Regulation of chromatin by histone modifications. *Cell Res.* **21**, 381–395 (2011).

272. Y. Zhao, B. A. Garcia, Comprehensive Catalog of Currently Documented Histone Modifications. *Cold Spring Harb. Perspect. Biol.* **7**, a025064 (2015).

273. M. C. Völker-Albert, A. Schmidt, I. Forne, A. Imhof, Analysis of Histone Modifications by Mass Spectrometry. *Curr. Protoc. Protein Sci.* **92**, e54 (2018).

274. C. Choudhary, B. T. Weinert, Y. Nishida, E. Verdin, M. Mann, The growing landscape of lysine acetylation links metabolism and cell signalling. *Nat. Rev. Mol. Cell Biol.* **15**, 536–550 (2014).

275. B. Egan, C.-C. Yuan, M. L. Craske, P. Labhart, G. D. Guler, D. Arnott, T. M. Maile, J. Busby, C. Henry, T. K. Kelly, C. A. Tindell, S. Jhunjhunwala, F. Zhao, C. Hatton, B. M. Bryant, M. Classon, P. Trojer, An Alternative Approach to ChIP-Seq Normalization Enables Detection of Genome-Wide

Changes in Histone H3 Lysine 27 Trimethylation upon EZH2 Inhibition. *PLoS One*. **11**, e0166438 (2016).

276. C. Schmidl, A. F. Rendeiro, N. C. Sheffield, C. Bock, ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nat. Methods*. **12**, 963–965 (2015).

277. M. Litt, Y. Qiu, S. Huang, Histone arginine methylations: their roles in chromatin dynamics and transcriptional regulation. *Biosci. Rep.* **29**, 131–141 (2009).

278. C. Choudhary, C. Kumar, F. Gnad, M. L. Nielsen, M. Rehman, T. C. Walther, J. V. Olsen, M. Mann, Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science*. **325**, 834–840 (2009).

279. J. Patel, R. R. Pathak, S. Mujtaba, The biology of lysine acetylation integrates transcriptional programming and metabolism. *Nutr. Metab.* . **8**, 12 (2011).

280. A. Sharda, R. V. Amnekar, A. Natu, Sukanya, S. Gupta, "Chapter 13 - Histone posttranslational modifications: Potential role in diagnosis, prognosis, and therapeutics of cancer" in *Prognostic Epigenetics*, S. Sharma, Ed. (Academic Press, 2019), vol. 15, pp. 351–373.

281. D. E. Sterner, S. L. Berger, Acetylation of histones and transcription-related factors. *Microbiol. Mol. Biol. Rev.* **64**, 435–459 (2000).

282. M. H. Kuo, C. D. Allis, Roles of histone acetyltransferases and deacetylases in gene regulation. *Bioessays*. **20**, 615–626 (1998).

283. B. C. Smith, J. M. Denu, Chemical mechanisms of histone lysine and arginine modifications. *Biochim. Biophys. Acta*. **1789**, 45–57 (2009).

284. H. Wapenaar, F. J. Dekker, Histone acetyltransferases: challenges in targeting bi-substrate enzymes. *Clin. Epigenetics*. **8**, 59 (2016).

285. X.-J. Yang, E. Seto, HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention. *Oncogene*. **26**, 5310–5318 (2007).

286. I. H. Lee, L. Cao, R. Mostoslavsky, D. B. Lombard, J. Liu, N. E. Bruns, M. Tsokos, F. W. Alt, T. Finkel, A role for the NAD-dependent deacetylase Sirt1 in the regulation of autophagy. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 3374–3379 (2008).

287. A. Brunet, L. B. Sweeney, J. F. Sturgill, K. F. Chua, P. L. Greer, Y. Lin, H. Tran, S. E. Ross, R. Mostoslavsky, H. Y. Cohen, L. S. Hu, H.-L. Cheng, M. P. Jedrychowski, S. P. Gygi, D. A. Sinclair, F. W. Alt, M. E. Greenberg, Stress-dependent regulation of FOXO transcription factors by the SIRT1 deacetylase. *Science*. **303**, 2011–2015 (2004).

288. E. L. Greer, Y. Shi, Histone methylation: a dynamic mark in health, disease and inheritance. *Nat. Rev. Genet.* **13**, 343–357 (2012).

289. A. Di Lorenzo, M. T. Bedford, Histone arginine methylation. *FEBS Lett.* **585**, 2024–2031 (2011).

290. A. Jambhekar, A. Dhall, Y. Shi, Roles and regulation of histone methylation in animal development. *Nat. Rev. Mol. Cell Biol.* **20**, 625–641 (2019).

291. R. S. Blanc, S. Richard, Arginine Methylation: The Coming of Age. *Mol. Cell.* **65**, 8–24 (2017).

292. D. Husmann, O. Gozani, Histone lysine methyltransferases in biology and disease. *Nat. Struct. Mol. Biol.* **26**, 880–889 (2019).

293. I. Pinheiro, R. Margueron, N. Shukeir, M. Eisold, C. Fritzsch, F. M. Richter, G. Mittler, C. Genoud, S. Goyama, M. Kurokawa, J. Son, D. Reinberg, M. Lachner, T. Jenuwein, Prdm3 and Prdm16 are H3K9me1 methyltransferases required for mammalian heterochromatin integrity. *Cell.* **150**, 948–960 (2012).

294. E. Metzger, S. Wang, S. Urban, D. Willmann, A. Schmidt, A. Offermann, A. Allen, M. Sum, N. Obier, F. Cottard, S. Ulferts, B.-T. Preca, B. Hermann, J. Maurer, H. Greschik, V. Hornung, O. Einsle, S. Perner, A. Imhof, M. Jung, R. Schüle, KMT9 monomethylates histone H4 lysine 12 and controls proliferation of prostate cancer cells. *Nat. Struct. Mol. Biol.* **26**, 361–371 (2019).

295. H. Mizukami, J.-D. Kim, S. Tabara, W. Lu, C. Kwon, M. Nakashima, A. Fukamizu, KDM5D-mediated H3K4 demethylation is required for sexually dimorphic gene expression in mouse embryonic fibroblasts. *J. Biochem.* **165**, 335–342 (2019).

296. E. Dimitrova, A. H. Turberfield, R. J. Klose, Histone demethylases in chromatin biology and beyond. *EMBO Rep.* **16**, 1620–1639 (2015).

297. D. Rossetto, N. Avvakumov, J. Côté, Histone phosphorylation: a chromatin modification involved in diverse nuclear events. *Epigenetics.* **7**, 1098–1108 (2012).

298. B. Ren, H. L. Tan, T. T. T. Nguyen, A. M. M. Sayed, Y. Li, Y.-K. Mok, H. Yang, E. S. Chen, Regulation of transcriptional silencing and chromodomain protein localization at centromeric heterochromatin by histone H3 tyrosine 41 phosphorylation in fission yeast. *Nucleic Acids Res.* **46**, 189–202 (2018).

299. M. Brehove, T. Wang, J. North, Y. Luo, S. J. Dreher, J. C. Shimko, J. J. Ottesen, K. Luger, M. G. Poirier, Histone core phosphorylation regulates DNA accessibility. *J. Biol. Chem.* **290**, 22612–22621 (2015).

300. X. Cao, W. Dang, "Histone Modification Changes During Aging" in *Epigenetics of Aging and Longevity* (Elsevier, 2018), pp. 309–328.

301. N. A. Watson, J. M. G. Higgins, "Histone Kinases and Phosphatases" in *Chromatin Signaling and Diseases* (Elsevier, 2016), pp. 75–94.

302. Z. Yang, M. He, J. Austin, J. Pfleger, M. Abdellatif, Histone H3K9 butyrylation is regulated by dietary fat and stress via an Acyl-CoA dehydrogenase short chain-dependent mechanism. *Mol Metab.* **53**, 101249 (2021).

303. Z. Kaczmarska, E. Ortega, A. Goudarzi, H. Huang, S. Kim, J. A. Márquez, Y. Zhao, S. Khochbin, D. Panne, Structure of p300 in complex with acyl-CoA variants. *Nat. Chem. Biol.* **13**, 21–29 (2017).

304. J. Garrity, J. G. Gardner, W. Hawse, C. Wolberger, J. C. Escalante-Semerena, N-lysine propionylation controls the activity of propionyl-CoA synthetase. *J. Biol. Chem.* **282**, 30239–30245 (2007).

305. B. Liu, Y. Lin, A. Darwanto, X. Song, G. Xu, K. Zhang, Identification and characterization of propionylation at histone H3 lysine 23 in mammalian cells. *J. Biol. Chem.* **284**, 32288–32295 (2009).

306. Y. Li, B. R. Sabari, T. Panchenko, H. Wen, D. Zhao, H. Guan, L. Wan, H. Huang, Z. Tang, Y.

Zhao, R. G. Roeder, X. Shi, C. D. Allis, H. Li, Molecular Coupling of Histone Crotonylation and Active Transcription by AF9 YEATS Domain. *Mol. Cell.* **62**, 181–193 (2016).

307. B. R. Sabari, Z. Tang, H. Huang, V. Yong-Gonzalez, H. Molina, H. E. Kong, L. Dai, M. Shimada, J. R. Cross, Y. Zhao, R. G. Roeder, C. D. Allis, Intracellular crotonyl-CoA stimulates transcription through p300-catalyzed histone crotonylation. *Mol. Cell.* **58**, 203–215 (2015).

308. X. Bao, Y. Wang, X. Li, X.-M. Li, Z. Liu, T. Yang, C. F. Wong, J. Zhang, Q. Hao, X. D. Li, Identification of "erasers" for lysine crotonylated histone marks using a chemical proteomics approach. *Elife.* **3** (2014), doi:10.7554/eLife.02999.

309. V. Dehennaut, D. Leprince, T. Lefebvre, O-GlcNAcylation, an Epigenetic Mark. Focus on the Histone Code, TET Family Proteins, and Polycomb Group Proteins. *Front. Endocrinol.* . **5**, 155 (2014).

310. R. Martinez-Zamudio, H. C. Ha, Histone ADP-ribosylation facilitates gene transcription by directly remodeling nucleosomes. *Mol. Cell. Biol.* **32**, 2490–2502 (2012).

311. Z. Xie, J. Dai, L. Dai, M. Tan, Z. Cheng, Y. Wu, J. D. Boeke, Y. Zhao, Lysine succinylation and lysine malonylation in histones. *Mol. Cell. Proteomics.* **11**, 100–107 (2012).

312. J. Liu, Y. Shangguan, D. Tang, Y. Dai, Histone succinylation and its function on the nucleosome. *J. Cell. Mol. Med.* **25**, 7101–7109 (2021).

313. V. Ginjala, K. Nacerddine, A. Kulkarni, J. Oza, S. J. Hill, M. Yao, E. Citterio, M. van Lohuizen, S. Ganesan, BMI1 is recruited to DNA breaks and contributes to DNA damage-induced H2A ubiquitination and repair. *Mol. Cell. Biol.* **31**, 1972–1982 (2011).

314. H. Kim, J. Chen, X. Yu, Ubiquitin-binding protein RAP80 mediates BRCA1-dependent DNA damage response. *Science.* **316**, 1202–1205 (2007).

315. J. A. Latham, R. J. Chosed, S. Wang, S. Y. R. Dent, Chromatin signaling to kinetochores: transregulation of Dam1 methylation by histone H2B ubiquitination. *Cell.* **146**, 709–719 (2011).

316. N. Minsky, E. Shema, Y. Field, M. Schuster, E. Segal, M. Oren, Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells. *Nat. Cell Biol.* **10**, 483–488 (2008).

317. S. Zaccara, R. J. Ries, S. R. Jaffrey, Reading, writing and erasing mRNA methylation. *Nat. Rev. Mol. Cell Biol.* **20**, 608–624 (2019).

318. R. F. Luco, M. Allo, I. E. Schor, A. R. Kornblihtt, T. Misteli, Epigenetics in alternative pre-mRNA splicing. *Cell.* **144**, 16–26 (2011).

319. J. M. Vaquerizas, S. K. Kummerfeld, S. A. Teichmann, N. M. Luscombe, A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.* **10**, 252–263 (2009).

320. K. D. Meyer, Y. Saletore, P. Zumbo, O. Elemento, C. E. Mason, S. R. Jaffrey, Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell.* **149**, 1635–1646 (2012).

321. D. Dominissini, S. Moshitch-Moshkovitz, S. Schwartz, M. Salmon-Divon, L. Ungar, S. Osenberg, K. Cesarkas, J. Jacob-Hirsch, N. Amariglio, M. Kupiec, R. Sorek, G. Rechavi, Topology of the

human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature*. **485**, 201–206 (2012).

322. F. Tuorto, R. Liebers, T. Musch, M. Schaefer, S. Hofmann, S. Kellner, M. Frye, M. Helm, G. Stoecklin, F. Lyko, RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. *Nat. Struct. Mol. Biol.* **19**, 900–905 (2012).

323. J. X. Cheng, L. Chen, Y. Li, A. Cloe, M. Yue, J. Wei, K. A. Watanabe, J. M. Shammo, J. Anastasi, Q. J. Shen, R. A. Larson, C. He, M. M. Le Beau, J. W. Vardiman, RNA cytosine methylation and methyltransferases mediate chromatin organization and 5-azacytidine response and resistance in leukaemia. *Nat. Commun.* **9**, 1163 (2018).

324. T. M. Carlile, M. F. Rojas-Duran, B. Zinshteyn, H. Shin, K. M. Bartoli, W. V. Gilbert, Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature*. **515**, 143–146 (2014).

325. N. M. Martinez, A. Su, M. C. Burns, J. K. Nussbacher, C. Schaening, S. Sathe, G. W. Yeo, W. V. Gilbert, Pseudouridine synthases modify human pre-mRNA co-transcriptionally and affect pre-mRNA processing. *Mol. Cell.* **82**, 645–659.e9 (2022).

326. W. Slotkin, K. Nishikura, Adenosine-to-inosine RNA editing and human disease. *Genome Med.* **5**, 105 (2013).

327. S. Stamm, S. Ben-Ari, I. Rafalska, Y. Tang, Z. Zhang, D. Toiber, T. A. Thanaraj, H. Soreq, Function of alternative splicing. *Gene*. **344**, 1–20 (2005).

328. W. Gilbert, Why genes in pieces? *Nature*. **271**, 501 (1978).

329. Y. Wang, J. Liu, B. O. Huang, Y.-M. Xu, J. Li, L.-F. Huang, J. Lin, J. Zhang, Q.-H. Min, W.-M. Yang, X.-Z. Wang, Mechanism of alternative splicing and its regulation. *Biomed Rep.* **3**, 152–158 (2015).

330. S. C. Bonnal, I. López-Oreja, J. Valcárcel, Roles and mechanisms of alternative splicing in cancer - implications for care. *Nat. Rev. Clin. Oncol.* **17**, 457–474 (2020).

331. E. R. Gibney, C. M. Nolan, Epigenetics and gene expression. *Heredity* . **105**, 4–13 (2010).

332. A. Vannini, P. Cramer, Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol. Cell.* **45**, 439–446 (2012).

333. J.-W. Wei, K. Huang, C. Yang, C.-S. Kang, Non-coding RNAs as regulators in epigenetics (Review). *Oncol. Rep.* **37**, 3–9 (2017).

334. C. P. Ponting, P. L. Oliver, W. Reik, Evolution and functions of long noncoding RNAs. *Cell.* **136**, 629–641 (2009).

335. K. Plath, S. Mlynarczyk-Evans, D. A. Nusinow, B. Panning, Xist RNA and the mechanism of X chromosome inactivation. *Annu. Rev. Genet.* **36**, 233–278 (2002).

336. P. Wang, J. Xu, Y. Wang, X. Cao, An interferon-independent lncRNA promotes viral replication by modulating cellular metabolism. *Science*. **358**, 1051–1055 (2017).

337. F. C. Beckedorff, A. C. Ayupe, R. Crocci-Souza, M. S. Amaral, H. I. Nakaya, D. T. Soltys, C. F. M. Menck, E. M. Reis, S. Verjovski-Almeida, The intronic long noncoding RNA ANRASSF1 recruits PRC2 to the RASSF1A promoter, reducing the expression of RASSF1A and increasing cell

proliferation. *PLoS Genet.* **9**, e1003705 (2013).

338. J. T. Lee, Epigenetic regulation by long noncoding RNAs. *Science.* **338**, 1435–1439 (2012).

339. Q. Yao, Y. Chen, X. Zhou, The roles of microRNAs in epigenetic regulation. *Curr. Opin. Chem. Biol.* **51**, 11–17 (2019).

340. A. Dharap, C. Pokrzywa, S. Murali, G. Pandi, R. Vemuganti, MicroRNA miR-324-3p induces promoter-mediated expression of RelA gene. *PLoS One.* **8**, e79467 (2013).

341. J. Zhang, W. Zhou, Y. Liu, T. Liu, C. Li, L. Wang, Oncogenic role of microRNA-532-5p in human colorectal cancer via targeting of the 5'UTR of. *Oncol. Lett.* **15**, 7215–7220 (2018).

342. J. J. Forman, A. Legesse-Miller, H. A. Coller, A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 14879–14884 (2008).

343. U. A. Ørom, F. C. Nielsen, A. H. Lund, MicroRNA-10a binds the 5'UTR of ribosomal protein mRNAs and enhances their translation. *Mol. Cell.* **30**, 460–471 (2008).

344. S. Vasudevan, J. A. Steitz, AU-rich-element-mediated upregulation of translation by FXR1 and Argonaute 2. *Cell.* **128**, 1105–1118 (2007).

345. M. Xiao, J. Li, W. Li, Y. Wang, F. Wu, Y. Xi, L. Zhang, C. Ding, H. Luo, Y. Li, L. Peng, L. Zhao, S. Peng, Y. Xiao, S. Dong, J. Cao, W. Yu, MicroRNAs activate gene transcription epigenetically as an enhancer trigger. *RNA Biol.* **14**, 1326–1334 (2017).

346. S. Liu, A. P. da Cunha, R. M. Rezende, R. Cialic, Z. Wei, L. Bry, L. E. Comstock, R. Gandhi, H. L. Weiner, The Host Shapes the Gut Microbiota via Fecal MicroRNA. *Cell Host Microbe.* **19**, 32–43 (2016).

347. D. M. Ozata, I. Gainetdinov, A. Zoch, D. O'Carroll, P. D. Zamore, PIWI-interacting RNAs: small RNAs with big functions. *Nat. Rev. Genet.* **20**, 89–108 (2019).

348. A. A. Aravin, R. Sachidanandam, D. Bourc'his, C. Schaefer, D. Pezic, K. F. Toth, T. Bestor, G. J. Hannon, A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol. Cell.* **31**, 785–799 (2008).

349. T. Kiss, Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell.* **109**, 145–148 (2002).

350. G. A. Stepanov, J. A. Filippova, A. B. Komissarov, E. V. Kuligina, V. A. Richter, D. V. Semenov, Regulatory role of small nucleolar RNAs in human diseases. *Biomed Res. Int.* **2015**, 206849 (2015).

351. N. H. Tolia, L. Joshua-Tor, Slicer and the argonautes. *Nat. Chem. Biol.* **3**, 36–43 (2007).

352. T. Sexton, E. Yaffe, E. Kenigsberg, F. Bantignies, B. Leblanc, M. Hoichman, H. Parrinello, A. Tanay, G. Cavalli, Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell.* **148**, 458–472 (2012).

353. B. Bonev, G. Cavalli, Organization and function of the 3D genome. *Nat. Rev. Genet.* **17**, 661–678 (2016).

354. A. Panigrahi, B. W. O'Malley, Mechanisms of enhancer action: the known and the unknown. *Genome Biol.* **22**, 108 (2021).

355. I. Miguel-Escalada, S. Bonàs-Guarch, I. Cebola, J. Ponsa-Cobas, J. Mendieta-Esteban, G. Atla, B. M. Javierre, D. M. Y. Rolando, I. Farabella, C. C. Morgan, J. García-Hurtado, A. Beucher, I. Morán, L. Pasquali, M. Ramos-Rodríguez, E. V. R. Appel, A. Linneberg, A. P. Gjesing, D. R. Witte, O. Pedersen, N. Grarup, P. Ravassard, D. Torrents, J. M. Mercader, L. Piemonti, T. Berney, E. J. P. de Koning, J. Kerr-Conte, F. Pattou, I. O. Fedko, L. Groop, I. Prokopenko, T. Hansen, M. A. Marti-Renom, P. Fraser, J. Ferrer, Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes. *Nat. Genet.* **51**, 1137–1148 (2019).

356. P. Li, L. Marshall, G. Oh, J. L. Jakubowski, D. Groot, Y. He, T. Wang, A. Petronis, V. Labrie, Epigenetic dysregulation of enhancers in neurons is associated with Alzheimer's disease pathology and cognitive symptoms. *Nat. Commun.* **10**, 2246 (2019).

357. A. B. Stergachis, S. Neph, R. Sandstrom, E. Haugen, A. P. Reynolds, M. Zhang, R. Byron, T. Canfield, S. Stelhing-Sun, K. Lee, R. E. Thurman, S. Vong, D. Bates, F. Neri, M. Diegel, E. Giste, D. Dunn, J. Vierstra, R. S. Hansen, A. K. Johnson, P. J. Sabo, M. S. Wilken, T. A. Reh, P. M. Treuting, R. Kaul, M. Groudine, M. A. Bender, E. Borenstein, J. A. Stamatoyannopoulos, Conservation of trans-acting circuitry during mammalian regulatory evolution. *Nature.* **515**, 365–370 (2014).

358. I. Jerkovic, G. Cavalli, Understanding 3D genome organization by multidisciplinary methods. *Nat. Rev. Mol. Cell Biol.* **22**, 511–528 (2021).

359. J. R. Wiśniewski, A. Zougman, N. Nagaraj, M. Mann, Universal sample preparation method for proteome analysis. *Nat. Methods.* **6**, 359–362 (2009).

360. G. J. Patti, O. Yanes, G. Siuzdak, Innovation: Metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* **13**, 263–269 (2012).

361. GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI, Biospecimen Collection Source Site—RPCI, Biospecimen Core Resource—VARI, Brain Bank Repository—University of Miami Brain Endowment Bank, Leidos Biomedical—Project Management, ELSI Study, Genome Browser Data Integration &Visualization—EBI, Genome Browser Data Integration &Visualization—UCSC Genomics Institute, University of California Santa Cruz, Lead analysts:, Laboratory, Data Analysis &Coordinating Center (LDACC):, NIH program management:, Biospecimen collection:, Pathology:, eQTL manuscript working group:, A. Battle, C. D. Brown, B. E. Engelhardt, S. B. Montgomery, Genetic effects on gene expression across human tissues. *Nature.* **550**, 204–213 (2017).

362. P. J. Park, ChIP-seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.* **10**, 669–680 (2009).

363. S. J. Clark, R. Argelaguet, C.-A. Kapourani, T. M. Stubbs, H. J. Lee, C. Alda-Catalinas, F. Krueger, G. Sanguinetti, G. Kelsey, J. C. Marioni, O. Stegle, W. Reik, scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nat. Commun.* **9**, 781 (2018).

364. G. Siuzdak, *The Expanding Role of Mass Spectrometry in Biotechnology* (MCC Press, 2003).

365. S. Banerjee, S. Mazumdar, Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *Int. J. Anal. Chem.* **2012**, 282574 (2012).

366. J. Capellades, A. Junza, S. Samino, J. S. Brunner, G. Schabbauer, M. Vinaixa, O. Yanes, Exploring the Use of Gas Chromatography Coupled to Chemical Ionization Mass Spectrometry (GC-CI-MS) for Stable Isotope Labeling in Metabolomics. *Anal. Chem.* **93**, 1242–1248 (2021).

367. M. Aichler, A. Walch, MALDI Imaging mass spectrometry: current frontiers and perspectives in pathology research and practice. *Lab. Invest.* **95**, 422–431 (2015).

368. E. Claude, E. A. Jones, S. D. Pringle, DESI Mass Spectrometry Imaging (MSI). *Methods Mol. Biol.* **1618**, 65–75 (2017).

369. L. J. Gamble, C. R. Anderton, Secondary Ion Mass Spectrometry Imaging of Tissues, Cells, and Microbial Systems. *Micros. Today*. **24**, 24–31 (2016).

370. R. Fischer, P. Bowness, B. M. Kessler, Two birds with one stone: doing metabolomics with your proteomics kit. *Proteomics*. **13**, 3371–3386 (2013).

371. G. L. Glish, D. J. Burinsky, Hybrid mass spectrometers for tandem mass spectrometry. *J. Am. Soc. Mass Spectrom.* **19**, 161–172 (2008).

372. C. Junot, F. Fenaille, B. Colsch, F. Bécher, High resolution mass spectrometry based techniques at the crossroads of metabolic pathways. *Mass Spectrom. Rev.* **33**, 471–500 (2014).

373. F. Meissner, J. Geddes-McAlister, M. Mann, M. Bantscheff, The emerging role of mass spectrometry-based proteomics in drug discovery. *Nat. Rev. Drug Discov.* **21**, 637–654 (2022).

374. D. S. Wishart, Emerging applications of metabolomics in drug discovery and precision medicine. *Nat. Rev. Drug Discov.* **15**, 473–484 (2016).

375. J. F. Xiao, B. Zhou, H. W. Ressom, Metabolite identification and quantitation in LC-MS/MS-based metabolomics. *Trends Analyt. Chem.* **32**, 1–14 (2012).

376. Y. Bian, R. Zheng, F. P. Bayer, C. Wong, Y.-C. Chang, C. Meng, D. P. Zolg, M. Reinecke, J. Zecha, S. Wiechmann, S. Heinzlmeir, J. Scherr, B. Hemmer, M. Baynham, A.-C. Gingras, O. Boychenko, B. Kuster, Robust, reproducible and quantitative analysis of thousands of proteomes by micro-flow LC-MS/MS. *Nat. Commun.* **11**, 157 (2020).

377. S. Keller, G. Jahreis, Determination of underivatised sterols and bile acid trimethyl silyl ether methyl esters by gas chromatography-mass spectrometry-single ion monitoring in faeces. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **813**, 199–207 (2004).

378. R. Wei, G. Li, A. B. Seymour, High-throughput and multiplexed LC/MS/MRM method for targeted metabolomics. *Anal. Chem.* **82**, 5527–5533 (2010).

379. P. Picotti, R. Aebersold, Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat. Methods*. **9**, 555–566 (2012).

380. A. C. Peterson, J. D. Russell, D. J. Bailey, M. S. Westphall, J. J. Coon, Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol. Cell. Proteomics*.

11, 1475–1488 (2012).

381. E. Defossez, J. Bourquin, S. von Reuss, S. Rasmann, G. Glauser, Eight key rules for successful data-dependent acquisition in mass spectrometry-based metabolomics. *Mass Spectrom. Rev.* **42**, 131–143 (2023).

382. B. C. Searle, K. E. Swearingen, C. A. Barnes, T. Schmidt, S. Gessulat, B. Küster, M. Wilhelm, Generating high quality libraries for DIA MS with empirically corrected peptide predictions. *Nat. Commun.* **11**, 1548 (2020).

383. A. Doerr, DIA mass spectrometry. *Nat. Methods.* **12**, 35–35 (2015).

384. L. Perez de Souza, S. Alseekh, F. Scossa, A. R. Fernie, Ultra-high-performance liquid chromatography high-resolution mass spectrometry variants for metabolomics research. *Nat. Methods.* **18**, 733–746 (2021).

385. W. Lu, B. D. Bennett, J. D. Rabinowitz, Analytical strategies for LC-MS-based targeted metabolomics. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **871**, 236–242 (2008).

386. N. L. Kuehnbaum, P. Britz-McKibbin, New advances in separation science for metabolomics: resolving chemical diversity in a post-genomic era. *Chem. Rev.* **113**, 2437–2468 (2013).

387. G. Baiges-Gaya, S. Iftimie, H. Castañé, E. Rodríguez-Tomàs, A. Jiménez-Franco, A. F. López-Azcona, A. Castro, J. Camps, J. Joven, Combining Semi-Targeted Metabolomics and Machine Learning to Identify Metabolic Alterations in the Serum and Urine of Hospitalized Patients with COVID-19. *Biomolecules.* **13** (2023), doi:10.3390/biom13010163.

388. W. Lu, X. Su, M. S. Klein, I. A. Lewis, O. Fiehn, J. D. Rabinowitz, Metabolite Measurement: Pitfalls to Avoid and Practices to Follow. *Annu. Rev. Biochem.* **86**, 277–304 (2017).

389. W. B. Dunn, D. Broadhurst, P. Begley, E. Zelena, S. Francis-McIntyre, N. Anderson, M. Brown, J. D. Knowles, A. Halsall, J. N. Haselden, A. W. Nicholls, I. D. Wilson, D. B. Kell, R. Goodacre, Human Serum Metabolome (HUSERMET) Consortium, Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* **6**, 1060–1083 (2011).

390. J. Zhou, Y. Yin, Strategies for large-scale targeted metabolomics quantification by liquid chromatography-mass spectrometry. *Analyst.* **141**, 6362–6373 (2016).

391. O. Yanes, R. Tautenhahn, G. J. Patti, G. Siuzdak, Expanding coverage of the metabolome for global metabolite profiling. *Anal. Chem.* **83**, 2152–2161 (2011).

392. D. Y. Lee, B. P. Bowen, T. R. Northen, Mass spectrometry-based metabolomics, analysis of metabolite-protein interactions, and imaging. *Biotechniques.* **49**, 557–565 (2010).

393. F. Bøgeskov Schmidt, A. M. Heskes, D. Thinagaran, B. Lindberg Møller, K. Jørgensen, B. A. Boughton, Mass Spectrometry Based Imaging of Labile Glucosides in Plants. *Front. Plant Sci.* **9**, 892 (2018).

394. L. Cui, H. Lu, Y. H. Lee, Challenges and emergent solutions for LC-MS/MS based untargeted metabolomics in diseases. *Mass Spectrom. Rev.* **37**, 772–792 (2018).

395. Z. Lai, O. Fiehn, Mass spectral fragmentation of trimethylsilylated small molecules. *Mass Spectrom.*

Rev. **37**, 245–257 (2018).

396. M. Katajamaa, M. Oresic, Data processing for mass spectrometry-based metabolomics. *J. Chromatogr. A.* **1158**, 318–328 (2007).

397. E. Want, Challenges in applying chemometrics to LC-MS-based global metabolite profile data. *Bioanalysis*. **1**, 805–819 (2009).

398. X. Domingo-Almenara, J. Brezmes, M. Vinaixa, S. Samino, N. Ramirez, M. Ramon-Krauel, C. Lerin, M. Díaz, L. Ibáñez, X. Correig, A. Perera-Lluna, O. Yanes, eRah: A Computational Tool Integrating Spectral Deconvolution and Alignment with Quantification and Identification of Metabolites in GC/MS-Based Metabolomics. *Anal. Chem.* **88**, 9821–9829 (2016).

399. A. Alonso, S. Marsal, A. Julià, Analytical methods in untargeted metabolomics: state of the art in 2015. *Front Bioeng Biotechnol.* **3**, 23 (2015).

400. M. Sugimoto, M. Kawakami, M. Robert, T. Soga, M. Tomita, Bioinformatics Tools for Mass Spectroscopy-Based Metabolomic Data Processing and Analysis. *Curr. Bioinform.* **7**, 96–108 (2012).

401. R. Smith, D. Ventura, J. T. Prince, LC-MS alignment in theory and practice: a comprehensive algorithmic review. *Brief. Bioinform.* **16**, 104–117 (2015).

402. M. Vinaixa, S. Samino, I. Saez, J. Duran, J. J. Guinovart, O. Yanes, A Guideline to Univariate Statistical Analysis for LC/MS-Based Untargeted Metabolomics-Derived Data. *Metabolites*. **2**, 775–795 (2012).

403. A. Chokkathukalam, D.-H. Kim, M. P. Barrett, R. Breitling, D. J. Creek, Stable isotope-labeling studies in metabolomics: new insights into structure and dynamics of metabolic networks. *Bioanalysis*. **6**, 511–524 (2014).

404. O. Senan, A. Aguilar-Mogas, M. Navarro, J. Capellades, L. Noon, D. Burks, O. Yanes, R. Guimerà, M. Sales-Pardo, CliqueMS: a computational tool for annotating in-source metabolite ions from LC-MS untargeted metabolomics data based on a coelution similarity network. *Bioinformatics*. **35**, 4089–4097 (2019).

405. X. Domingo-Almenara, J. R. Montenegro-Burke, H. P. Benton, G. Siuzdak, Annotation: A Computational Solution for Streamlining Metabolomics Analysis. *Anal. Chem.* **90**, 480–489 (2018).

406. C. D. Broeckling, F. A. Afsar, S. Neumann, A. Ben-Hur, J. E. Prenni, RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. *Anal. Chem.* **86**, 6812–6817 (2014).

407. A. Mastrangelo, A. Ferrarini, F. Rey-Stolle, A. García, C. Barbas, From sample treatment to biomarker discovery: A tutorial for untargeted metabolomics based on GC-(EI)-Q-MS. *Anal. Chim. Acta.* **900**, 21–35 (2015).

408. M. Vinaixa, E. L. Schymanski, S. Neumann, M. Navarro, R. M. Salek, O. Yanes, Mass spectral databases for LC/MS- and GC/MS-based metabolomics: State of the field and future prospects. *Trends Analyt. Chem.* **78**, 23–35 (2016).

409. T. Kind, H. Tsugawa, T. Cajka, Y. Ma, Z. Lai, S. S. Mehta, G. Wohlgemuth, D. K. Barupal, M. R. Showalter, M. Arita, O. Fiehn, Identification of small molecules using accurate mass MS/MS search. *Mass Spectrom. Rev.* **37**, 513–532 (2018).

410. A. Srivastava, G. M. Kowalski, D. L. Callahan, P. J. Meikle, D. J. Creek, Strategies for Extending Metabolomics Studies with Stable Isotope Labelling and Fluxomics. *Metabolites*. **6** (2016), doi:10.3390/metabo6040032.

411. J. Capellades, M. Navarro, S. Samino, M. Garcia-Ramirez, C. Hernandez, R. Simo, M. Vinaixa, O. Yanes, geoRge: A Computational Tool To Detect the Presence of Stable Isotope Labeling in LC/MS-Based Untargeted Metabolomics. *Anal. Chem.* **88**, 621–628 (2016).

412. C. Jang, L. Chen, J. D. Rabinowitz, Metabolomics and Isotope Tracing. *Cell.* **173**, 822–837 (2018).

413. N. Zamboni, S.-M. Fendt, M. Rühl, U. Sauer, (13)C-based metabolic flux analysis. *Nat. Protoc.* **4**, 878–892 (2009).

414. M. Cascante, S. Marin, Metabolomics and fluxomics approaches. *Essays Biochem.* **45**, 67–81 (2008).

415. R. Aebersold, M. Mann, Mass spectrometry-based proteomics. *Nature*. **422**, 198–207 (2003).

416. A. D. Catherman, O. S. Skinner, N. L. Kelleher, Top Down proteomics: facts and perspectives. *Biochem. Biophys. Res. Commun.* **445**, 683–693 (2014).

417. A. F. M. Altelaar, J. Munoz, A. J. R. Heck, Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat. Rev. Genet.* **14**, 35–48 (2013).

418. V. Vidova, Z. Spacil, A review on mass spectrometry-based quantitative proteomics: Targeted and data independent acquisition. *Anal. Chim. Acta*. **964**, 7–23 (2017).

419. F. Sanger, S. Nicklen, A. R. Coulson, DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467 (1977).

420. E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, Y. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de

la Bastide, N. Dedhia, H. Blöcker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowki, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, Y. J. Chen, J. Szustakowki, International Human Genome Sequencing Consortium, Initial sequencing and analysis of the human genome. *Nature*. **409**, 860–921 (2001).

421. J. Shendure, S. Balasubramanian, G. M. Church, W. Gilbert, J. Rogers, J. A. Schloss, R. H. Waterston, DNA sequencing at 40: past, present and future. *Nature*. **550**, 345–353 (2017).

422. S. Xie, A. W.-S. Leung, Z. Zheng, D. Zhang, C. Xiao, R. Luo, M. Luo, S. Zhang, Applications and potentials of nanopore sequencing in the (epi)genome and (epi)transcriptome era. *Innovation (Camb)*. **2**, 100153 (2021).

423. A. Rhoads, K. F. Au, PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics*. **13**, 278–289 (2015).

424. J. Eid, A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman, A. Bibillo, K. Bjornson, B. Chaudhuri, F. Christians, R. Cicero, S. Clark, R. Dalal, A. Dewinter, J. Dixon, M. Foquet, A. Gaertner, P. Hardenbol, C. Heiner, K. Hester, D. Holden, G. Kearns, X. Kong, R. Kuse, Y. Lacroix, S. Lin, P. Lundquist, C. Ma, P. Marks, M. Maxham, D. Murphy, I. Park, T. Pham, M. Phillips, J. Roy, R. Sebra, G. Shen, J. Sorenson, A. Tomaney, K. Travers, M. Trulson, J. Vieceli, J. Wegener, D. Wu, A. Yang, D. Zaccarin, P. Zhao, F. Zhong, J. Korlach, S. Turner, Real-time DNA sequencing from single polymerase molecules. *Science*. **323**, 133–138 (2009).

425. A. Conesa, P. Madrigal, S. Tarazona, D. Gomez-Cabrero, A. Cervera, A. McPherson, M. W. Szcześniak, D. J. Gaffney, L. L. Elo, X. Zhang, A. Mortazavi, A survey of best practices for RNA-seq data analysis. *Genome Biol.* **17**, 13 (2016).

426. S. M. Berget, C. Moore, P. A. Sharp, Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 3171–3175 (1977).

427. M. Melé, K. Mattioli, W. Mallard, D. M. Shechner, C. Gerhardinger, J. L. Rinn, Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs. *Genome Res.* **27**, 27–37 (2017).

428. F. Mitelman, B. Johansson, F. Mertens, Fusion genes and rearranged genes as a linear function of chromosome aberrations in cancer. *Nat. Genet.* **36**, 331–334 (2004).

429. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods*. **9**, 357–359 (2012).

430. D. Kim, J. M. Paggi, C. Park, C. Bennett, S. L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).

431. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R.

Gingeras, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. **29**, 15–21 (2013).

432. S. Anders, P. T. Pyl, W. Huber, HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*. **31**, 166–169 (2015).

433. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).

434. R. Patro, G. Duggal, M. I. Love, R. A. Irizarry, C. Kingsford, Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods*. **14**, 417–419 (2017).

435. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

436. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. **26**, 139–140 (2010).

437. H. Zhu, G. Wang, J. Qian, Transcription factors as readers and effectors of DNA methylation. *Nat. Rev. Genet.* **17**, 551–565 (2016).

438. H. Guo, B. Hu, L. Yan, J. Yong, Y. Wu, Y. Gao, F. Guo, Y. Hou, X. Fan, J. Dong, X. Wang, X. Zhu, J. Yan, Y. Wei, H. Jin, W. Zhang, L. Wen, F. Tang, J. Qiao, DNA methylation and chromatin accessibility profiling of mouse and human fetal germ cells. *Cell Res.* **27**, 165–183 (2017).

439. H. M. O'Hagan, W. Wang, S. Sen, C. Destefano Shields, S. S. Lee, Y. W. Zhang, E. G. Clements, Y. Cai, L. Van Neste, H. Easwaran, R. A. Casero, C. L. Sears, S. B. Baylin, Oxidative damage targets complexes containing DNA methyltransferases, SIRT1, and polycomb members to promoter CpG Islands. *Cancer Cell.* **20**, 606–619 (2011).

440. M. Jung, G. P. Pfeifer, Aging and DNA methylation. *BMC Biol.* **13**, 7 (2015).

441. A. Nishiyama, M. Nakanishi, Navigating the DNA methylation landscape of cancer. *Trends Genet.* **37**, 1012–1027 (2021).

442. L. Liu, T. van Groen, I. Kadish, T. O. Tollefsbol, DNA methylation impacts on learning and memory in aging. *Neurobiol. Aging.* **30**, 549–560 (2009).

443. M. Jackson, A. Krassowska, N. Gilbert, T. Chevassut, L. Forrester, J. Ansell, B. Ramsahoye, Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells. *Mol. Cell. Biol.* **24**, 8862–8871 (2004).

444. M. Ehrlich, DNA hypomethylation in cancer cells. *Epigenomics*. **1**, 239–259 (2009).

445. V. Lloréns-Rico, J. A. Simcock, G. R. B. Huys, J. Raes, Single-cell approaches in human microbiome research. *Cell.* **185**, 2725–2738 (2022).

446. W. Zheng, S. Zhao, Y. Yin, H. Zhang, D. M. Needham, E. D. Evans, C. L. Dai, P. J. Lu, E. J. Alm, D. A. Weitz, High-throughput, single-microbe genomics with strain resolution, applied to a human gut microbiome. *Science*. **376**, eabm1483 (2022).

447. P. J. Turnbaugh, C. Quince, J. J. Faith, A. C. McHardy, T. Yatsunenko, F. Niazi, J. Affourtit, M. Egholm, B. Henrissat, R. Knight, J. I. Gordon, Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. *Proc. Natl. Acad. Sci. U. S. A.* **107**,

7503–7508 (2010).

448. D. S. Clausen, A. D. Willis, Evaluating replicability in microbiome data. *Biostatistics*. **23**, 1099–1114 (2022).

449. R. Knight, A. Vrbanac, B. C. Taylor, A. Aksenov, C. Callewaert, J. Debelius, A. Gonzalez, T. Kosciolek, L.-I. McCall, D. McDonald, A. V. Melnik, J. T. Morton, J. Navas, R. A. Quinn, J. G. Sanders, A. D. Swafford, L. R. Thompson, A. Tripathi, Z. Z. Xu, J. R. Zaneveld, Q. Zhu, J. G. Caporaso, P. C. Dorrestein, Best practices for analysing microbiomes. *Nat. Rev. Microbiol.* **16**, 410–422 (2018).

450. V. Pawlowsky-Glahn, J. J. Egozcue, R. Tolosana-Delgado, *Modeling and Analysis of Compositional Data* (John Wiley & Sons, 2015).

451. G. B. Gloor, J. M. Macklaim, V. Pawlowsky-Glahn, J. J. Egozcue, Microbiome Datasets Are Compositional: And This Is Not Optional. *Front. Microbiol.* **8**, 2224 (2017).

452. B. J. Callahan, P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, S. P. Holmes, DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods*. **13**, 581–583 (2016).

453. P. J. McMurdie, S. Holmes, phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*. **8**, e61217 (2013).

454. E. Bolyen, J. R. Rideout, M. R. Dillon, N. A. Bokulich, C. C. Abnet, G. A. Al-Ghalith, H. Alexander, E. J. Alm, M. Arumugam, F. Asnicar, Y. Bai, J. E. Bisanz, K. Bittinger, A. Brejnrod, C. J. Brislawn, C. T. Brown, B. J. Callahan, A. M. Caraballo-Rodríguez, J. Chase, E. K. Cope, R. Da Silva, C. Diener, P. C. Dorrestein, G. M. Douglas, D. M. Durall, C. Duvallet, C. F. Edwardson, M. Ernst, M. Estaki, J. Fouquier, J. M. Gauglitz, S. M. Gibbons, D. L. Gibson, A. Gonzalez, K. Gorlick, J. Guo, B. Hillmann, S. Holmes, H. Holste, C. Huttenhower, G. A. Huttley, S. Janssen, A. K. Jarmusch, L. Jiang, B. D. Kaehler, K. B. Kang, C. R. Keefe, P. Keim, S. T. Kelley, D. Knights, I. Koester, T. Kosciolek, J. Kreps, M. G. I. Langille, J. Lee, R. Ley, Y.-X. Liu, E. Loftfield, C. Lozupone, M. Maher, C. Marotz, B. D. Martin, D. McDonald, L. J. McIver, A. V. Melnik, J. L. Metcalf, S. C. Morgan, J. T. Morton, A. T. Naimey, J. A. Navas-Molina, L. F. Nothias, S. B. Orchanian, T. Pearson, S. L. Peoples, D. Petras, M. L. Preuss, E. Pruesse, L. B. Rasmussen, A. Rivers, M. S. Robeson 2nd, P. Rosenthal, N. Segata, M. Shaffer, A. Shiffer, R. Sinha, S. J. Song, J. R. Spear, A. D. Swafford, L. R. Thompson, P. J. Torres, P. Trinh, A. Tripathi, P. J. Turnbaugh, S. Ul-Hasan, J. J. J. van der Hooft, F. Vargas, Y. Vázquez-Baeza, E. Vogtmann, M. von Hippel, W. Walters, Y. Wan, M. Wang, J. Warren, K. C. Weber, C. H. D. Williamson, A. D. Willis, Z. Z. Xu, J. R. Zaneveld, Y. Zhang, Q. Zhu, R. Knight, J. G. Caporaso, Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).

455. E. Kopylova, L. Noé, H. Touzet, SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*. **28**, 3211–3217 (2012).

456. A. Blanco-Míguez, F. Beghini, F. Cumbo, L. J. McIver, K. N. Thompson, M. Zolfo, P. Manghi, L. Dubois, K. D. Huang, A. M. Thomas, W. A. Nickols, G. Piccinno, E. Piperni, M. Punčochář, M. Valles-Colomer, A. Tett, F. Giordano, R. Davies, J. Wolf, S. E. Berry, T. D. Spector, E. A. Franzosa, E. Pasolli, F. Asnicar, C. Huttenhower, N. Segata, Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat. Biotechnol.* (2023), doi:10.1038/s41587-023-01688-w.

457. E. A. Franzosa, L. J. McIver, G. Rahnavard, L. R. Thompson, M. Schirmer, G. Weingart, K. S.

Lipson, R. Knight, J. G. Caporaso, N. Segata, C. Huttenhower, Species-level functional profiling of metagenomes and metatranscriptomes. *Nat. Methods*. **15**, 962–968 (2018).

458. D. H. Huson, S. Beier, I. Flade, A. Górska, M. El-Hadidi, S. Mitra, H.-J. Ruscheweyh, R. Tappu, MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Comput. Biol.* **12**, e1004957 (2016).

459. J. Thorsen, A. Brejnrod, M. Mortensen, M. A. Rasmussen, J. Stokholm, W. A. Al-Soud, S. Sørensen, H. Bisgaard, J. Waage, Large-scale benchmarking reveals false discoveries and count transformation sensitivity in 16S rRNA gene amplicon data analysis methods used in microbiome studies. *Microbiome*. **4**, 62 (2016).

460. S. Hawinkel, F. Mattiello, L. Bijnens, O. Thas, A broken promise: microbiome differential abundance methods do not control the false discovery rate. *Brief. Bioinform.* **20**, 210–221 (2019).

461. E. C. Pielou, The measurement of diversity in different types of biological collections. *J. Theor. Biol.* **13**, 131–144 (1966).

462. H. Shimadzu, On species richness and rarefaction: size- and coverage-based techniques quantify different characteristics of richness change in biodiversity. *J. Math. Biol.* **77**, 1363–1381 (2018).

463. C. E. Shannon, A mathematical theory of communication. *Bell Syst. tech. j.* **27**, 379–423 (1948).

464. R. H. Whittaker, Vegetation of the siskiyou mountains, Oregon and California. *Ecol. Monogr.* **30**, 279–338 (1960).

465. I. Subramanian, S. Verma, S. Kumar, A. Jere, K. Anamika, Multi-omics Data Integration, Interpretation, and Its Application. *Bioinform. Biol. Insights*. **14**, 1177932219899051 (2020).

466. M. Bersanelli, E. Mosca, D. Remondini, E. Giampieri, C. Sala, G. Castellani, L. Milanesi, Methods for the integration of multi-omics data: mathematical aspects. *BMC Bioinformatics*. **17 Suppl 2**, 15 (2016).

467. R. Argelaguet, B. Velten, D. Arnol, S. Dietrich, T. Zenz, J. C. Marioni, F. Buettner, W. Huber, O. Stegle, Multi-Omics Factor Analysis-a framework for unsupervised integration of multi-omics data sets. *Mol. Syst. Biol.* **14**, e8124 (2018).

468. R. Argelaguet, D. Arnol, D. Bredikhin, Y. Deloro, B. Velten, J. C. Marioni, O. Stegle, MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol.* **21**, 111 (2020).

469. F. Rohart, B. Gautier, A. Singh, K.-A. Lê Cao, mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Comput. Biol.* **13**, e1005752 (2017).

470. J. A. Seoane, I. N. M. Day, T. R. Gaunt, C. Campbell, A pathway-based data integration framework for prediction of disease progression. *Bioinformatics*. **30**, 838–845 (2014).

471. C. Gao, J. Liu, A. R. Kriebel, S. Preissl, C. Luo, R. Castanon, J. Sandoval, A. Rivkin, J. R. Nery, M. M. Behrens, J. R. Ecker, B. Ren, J. D. Welch, Iterative single-cell multi-omic integration using online learning. *Nat. Biotechnol.* **39**, 1000–1007 (2021).

472. N. Rappoport, R. Shamir, NEMO: cancer subtyping by integration of partial multi-omic data. *Bioinformatics*. **35**, 3348–3356 (2019).

473. R. Shen, A. B. Olshen, M. Ladanyi, Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinformatics.* **25**, 2906–2912 (2009).

474. P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* **9**, 559 (2008).

475. R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Y. H. Yang, J. Zhang, Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).

476. L. Cantini, P. Zakeri, C. Hernandez, A. Naldi, D. Thieffry, E. Remy, A. Baudot, Benchmarking joint multi-omics dimensionality reduction approaches for the study of cancer. *Nat. Commun.* **12**, 124 (2021).

477. C. Clark, L. Dayon, M. Masoodi, G. L. Bowman, J. Popp, An integrative multi-omics approach reveals new central nervous system pathway alterations in Alzheimer's disease. *Alzheimers. Res. Ther.* **13**, 71 (2021).

478. L. Mangiante, N. Alcala, A. Sexton-Oates, A. Di Genova, A. Gonzalez-Perez, A. Khandekar, E. N. Bergstrom, J. Kim, X. Liu, R. Blazquez-Encinas, C. Giacobi, N. Le Stang, S. Boyault, C. Cuenin, S. Tabone-Eglinger, F. Damiola, C. Voegele, M. Ardin, M.-C. Michallet, L. Soudade, T. M. Delhomme, A. Poret, M. Brevet, M.-C. Copin, S. Giusiano-Courcambeck, D. Damotte, C. Girard, V. Hofman, P. Hofman, J. Mouroux, C. Cohen, S. Lacomme, J. Mazieres, V. T. de Montpreville, C. Perrin, G. Planchard, N. Rousseau, I. Rouquette, C. Sagan, A. Scherpereel, F. Thivolet, J.-M. Vignaud, D. Jean, A. G. S. Ilg, R. Olaso, V. Meyer, A. Boland-Auge, J.-F. Deleuze, J. Altmuller, P. Nuernberg, A. Ibáñez-Costa, J. P. Castaño, S. Lantuejoul, A. Ghantous, C. Maussion, P. Courtiol, H. Hernandez-Vargas, C. Caux, N. Girard, N. Lopez-Bigas, L. B. Alexandrov, F. Galateau-Salle, M. Foll, L. Fernandez-Cuesta, Multiomic analysis of malignant pleural mesothelioma identifies molecular axes and specialized tumor profiles driving intertumor heterogeneity. *Nat. Genet.* **55**, 607–618 (2023).

479. F. Hugenholtz, W. M. de Vos, Mouse models for human intestinal microbiota research: a critical evaluation. *Cell. Mol. Life Sci.* **75**, 149–160 (2018).

480. H. Kwon, J. M. Schafer, N.-J. Song, S. Kaneko, A. Li, T. Xiao, A. Ma, C. Allen, K. Das, L. Zhou, B. Riesenberg, Y. Chang, P. Weltge, M. Velegraki, D. Y. Oh, L. Fong, Q. Ma, D. Sundi, D. Chung, X. Li, Z. Li, Androgen conspires with the CD8 T cell exhaustion program and contributes to sex bias in cancer. *Sci Immunol.* **7**, eabq2630 (2022).

481. K.-A. Kim, W. Gu, I.-A. Lee, E.-H. Joh, D.-H. Kim, High fat diet-induced gut microbiota exacerbates inflammation and obesity in mice via the TLR4 signaling pathway. *PLoS One.* **7**, e47713 (2012).

482. E. A. Murphy, K. T. Velazquez, K. M. Herbert, Influence of high-fat diet on gut microbiota: a driving force for chronic disease risk. *Curr. Opin. Clin. Nutr. Metab. Care.* **18**, 515–520 (2015).

483. F. Bäckhed, H. Ding, T. Wang, L. V. Hooper, G. Y. Koh, A. Nagy, C. F. Semenkovich, J. I. Gordon, The gut microbiota as an environmental factor that regulates fat storage. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 15718–15723 (2004).

484. C. Kong, R. Gao, X. Yan, L. Huang, H. Qin, Probiotics improve gut microbiota dysbiosis in obese mice fed a high-fat or high-sucrose diet. *Nutrition*. **60**, 175–184 (2019).

485. T. Chunchai, W. Thunapong, S. Yasom, K. Wanchai, S. Eaimworawuthikul, G. Metzler, A. Lungkaphin, A. Pongchaidecha, S. Sirilun, C. Chaiyasut, W. Pratchayasakul, P. Thiennimitr, N. Chattipakorn, S. C. Chattipakorn, Decreased microglial activation through gut-brain axis by prebiotics, probiotics, or synbiotics effectively restored cognitive function in obese-insulin resistant rats. *J. Neuroinflammation*. **15**, 11 (2018).

486. D. Zhou, Q. Pan, F. Shen, H.-X. Cao, W.-J. Ding, Y.-W. Chen, J.-G. Fan, Total fecal microbiota transplantation alleviates high-fat diet-induced steatohepatitis in mice via beneficial regulation of gut microbiota. *Sci. Rep.* **7**, 1529 (2017).

487. J. Torres, J. Hu, A. Seki, C. Eisele, N. Nair, R. Huang, L. Tarassishin, B. Jharap, J. Cote-Daigneault, Q. Mao, I. Mogno, G. J. Britton, M. Uzzan, C.-L. Chen, A. Kornbluth, J. George, P. Legnani, E. Maser, H. Loudon, J. Stone, M. Dubinsky, J. J. Faith, J. C. Clemente, S. Mehandru, J.-F. Colombel, I. Peter, Infants born to mothers with IBD present with altered gut microbiome that transfers abnormalities of the adaptive immune system to germ-free mice. *Gut*. **69**, 42–51 (2020).

488. G. Chevalier, E. Siopi, L. Guenin-Macé, M. Pascal, T. Laval, A. Rifflet, I. G. Boneca, C. Demangel, B. Colsch, A. Pruvost, E. Chu-Van, A. Messager, F. Leulier, G. Lepousez, G. Eberl, P.-M. Lledo, Effect of gut microbiota on depressive-like behaviors in mice is mediated by the endocannabinoid system. *Nat. Commun.* **11**, 6363 (2020).

489. R. Jumpertz, D. S. Le, P. J. Turnbaugh, C. Trinidad, C. Bogardus, J. I. Gordon, J. Krakoff, Energy-balance studies reveal associations between gut microbes, caloric load, and nutrient absorption in humans. *Am. J. Clin. Nutr.* **94**, 58–65 (2011).

490. M. Zimmermann, M. Zimmermann-Kogadeeva, R. Wegmann, A. L. Goodman, Mapping human microbiome drug metabolism by gut bacteria and their genes. *Nature*. **570**, 462–467 (2019).

491. E. A. Kennedy, K. Y. King, M. T. Baldridge, Mouse Microbiota Models: Comparing Germ-Free Mice and Antibiotics Treatment as Tools for Modifying Gut Bacteria. *Front. Physiol.* **9**, 1534 (2018).

492. Z.-L. Lai, C.-H. Tseng, H. J. Ho, C. K. Y. Cheung, J.-Y. Lin, Y.-J. Chen, F.-C. Cheng, Y.-C. Hsu, J.-T. Lin, E. M. El-Omar, C.-Y. Wu, Fecal microbiota transplantation confers beneficial metabolic effects of diet and exercise on diet-induced obese mice. *Sci. Rep.* **8**, 15625 (2018).

493. V. K. Ridaura, J. J. Faith, F. E. Rey, J. Cheng, A. E. Duncan, A. L. Kau, N. W. Griffin, V. Lombard, B. Henrissat, J. R. Bain, M. J. Muehlbauer, O. Ilkayeva, C. F. Semenkovich, K. Funai, D. K. Hayashi, B. J. Lyle, M. C. Martini, L. K. Ursell, J. C. Clemente, W. Van Treuren, W. A. Walters, R. Knight, C. B. Newgard, A. C. Heath, J. I. Gordon, Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science*. **341**, 1241214 (2013).

494. M. N. Quraishi, M. Widlak, N. Bhala, D. Moore, M. Price, N. Sharma, T. H. Iqbal, Systematic review with meta-analysis: the efficacy of faecal microbiota transplantation for the treatment of recurrent and refractory Clostridium difficile infection. *Aliment. Pharmacol. Ther.* **46**, 479–493 (2017).

495. Y. Chen, Z. Xueying, C. Jiaqu, C. Qiyi, Q. Huanlong, L. Ning, D. Yasong, Z. Xiaoxin, Y. Rong, L. Jubao, L. Xiaoqiong, M. Chunlian, W. Yu, C. Shidong, K. Guifang, Z. Dongmei, F. Shuanfeng, Z. Xujing, Y. Binrang, W. Yanxia, L. Ling, Y. Song, Z. Xiang, Z. Beihua, J. Lin, J. Hong, FTACMT

Working Group, FTACMT study protocol: a multicentre, double-blind, randomised, placebo-controlled trial of faecal microbiota transplantation for autism spectrum disorder. *BMJ Open*. **12**, e051613 (2022).

496. F. Zhang, N. Qi, Y. Zeng, M. Bao, Y. Chen, J. Liao, L. Wei, D. Cao, S. Huang, Q. Luo, Y. Jiang, Z. Mo, The Endogenous Alterations of the Gut Microbiota and Feces Metabolites Alleviate Oxidative Damage in the Brain of LanCL1 Knockout Mice. *Front. Microbiol.* **11**, 557342 (2020).

497. C. Gubert, G. Kong, V. Uzungil, A. M. Zeleznikow-Johnston, E. L. Burrows, T. Renoir, A. J. Hannan, Microbiome Profiling Reveals Gut Dysbiosis in the Metabotropic Glutamate Receptor 5 Knockout Mouse Model of Schizophrenia. *Front Cell Dev Biol.* **8**, 582320 (2020).

498. S. J. Robertson, P. Lemire, H. Maughan, A. Goethel, W. Turpin, L. Bedrani, D. S. Guttman, K. Croitoru, S. E. Girardin, D. J. Philpott, Comparison of Co-housing and Littermate Methods for Microbiota Standardization in Mouse Models. *Cell Rep.* **27**, 1910–1919.e2 (2019).

499. R. Lundberg, M. F. Toft, B. August, A. K. Hansen, C. H. F. Hansen, Antibiotic-treated versus germ-free rodents for microbiota transplantation studies. *Gut Microbes*. **7**, 68–74 (2016).

500. K. M. Ng, A. Aranda-Díaz, C. Tropini, M. R. Frankel, W. Van Treuren, C. T. O'Loughlin, B. D. Merrill, F. B. Yu, K. M. Pruss, R. A. Oliveira, S. K. Higginbottom, N. F. Neff, M. A. Fischbach, K. B. Xavier, J. L. Sonnenburg, K. C. Huang, Recovery of the Gut Microbiota after Antibiotics Depends on Host Diet, Community Context, and Environmental Reservoirs. *Cell Host Microbe*. **26**, 650–665.e4 (2019).

501. L. K. Ursell, J. L. Metcalf, L. W. Parfrey, R. Knight, Defining the human microbiome. *Nutr. Rev.* **70 Suppl 1**, S38–44 (2012).

502. D. C. Savage, Microbial ecology of the gastrointestinal tract. *Annu. Rev. Microbiol.* **31**, 107–133 (1977).

503. R. Sender, S. Fuchs, R. Milo, Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans. *Cell*. **164**, 337–340 (2016).

504. T. R. Sampson, S. K. Mazmanian, Control of brain development, function, and behavior by the microbiome. *Cell Host Microbe*. **17**, 565–576 (2015).

505. J. M. Yano, K. Yu, G. P. Donaldson, G. G. Shastri, P. Ann, L. Ma, C. R. Nagler, R. F. Ismagilov, S. K. Mazmanian, E. Y. Hsiao, Indigenous bacteria from the gut microbiota regulate host serotonin biosynthesis. *Cell*. **161**, 264–276 (2015).

506. P. Strandwitz, K. H. Kim, D. Terekhova, J. K. Liu, A. Sharma, J. Levering, D. McDonald, D. Dietrich, T. R. Ramadhar, A. Lekbua, N. Mroue, C. Liston, E. J. Stewart, M. J. Dubin, K. Zengler, R. Knight, J. A. Gilbert, J. Clardy, K. Lewis, GABA-modulating bacteria of the human gut microbiota. *Nat Microbiol.* **4**, 396–403 (2019).

507. A. L. Kau, P. P. Ahern, N. W. Griffin, A. L. Goodman, J. I. Gordon, Human nutrition, the gut microbiome and the immune system. *Nature*. **474**, 327–336 (2011).

508. D. Reijnders, G. H. Goossens, G. D. A. Hermes, E. P. J. G. Neis, C. M. van der Beek, J. Most, J. J. Holst, K. Lenaerts, R. S. Kootte, M. Nieuwdorp, A. K. Groen, S. W. M. Olde Damink, M. V. Boekschoten, H. Smidt, E. G. Zoetendal, C. H. C. Dejong, E. E. Blaak, Effects of Gut Microbiota Manipulation by Antibiotics on Host Metabolism in Obese Humans: A Randomized

Double-Blind Placebo-Controlled Trial. *Cell Metab.* **24**, 63–74 (2016).

509. H. Chu, S. K. Mazmanian, Innate immune recognition of the microbiota promotes host-microbial symbiosis. *Nat. Immunol.* **14**, 668–675 (2013).

510. K. Watanabe, C. A. Gilchrist, M. J. Uddin, S. L. Burgess, M. M. Abhyankar, S. N. Moonah, Z. Noor, J. R. Donowitz, B. N. Schneider, T. Arju, E. Ahmed, M. Kabir, M. Alam, R. Haque, P. Pramoonjago, B. Mehrad, W. A. Petri Jr, Microbiome-mediated neutrophil recruitment via CXCR2 and protection from amebic colitis. *PLoS Pathog.* **13**, e1006513 (2017).

511. P. J. Turnbaugh, F. Bäckhed, L. Fulton, J. I. Gordon, Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host Microbe.* **3**, 213–223 (2008).

512. C. De Filippo, D. Cavalieri, M. Di Paola, M. Ramazzotti, J. B. Poullet, S. Massart, S. Collini, G. Pieraccini, P. Lionetti, Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 14691–14696 (2010).

513. B. D. Muegge, J. Kuczynski, D. Knights, J. C. Clemente, A. González, L. Fontana, B. Henrissat, R. Knight, J. I. Gordon, Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science.* **332**, 970–974 (2011).

514. G. D. Wu, J. Chen, C. Hoffmann, K. Bittinger, Y.-Y. Chen, S. A. Keilbaugh, M. Bewtra, D. Knights, W. A. Walters, R. Knight, R. Sinha, E. Gilroy, K. Gupta, R. Baldassano, L. Nessel, H. Li, F. D. Bushman, J. D. Lewis, Linking long-term dietary patterns with gut microbial enterotypes. *Science.* **334**, 105–108 (2011).

515. C. Menni, M. A. Jackson, T. Pallister, C. J. Steves, T. D. Spector, A. M. Valdes, Gut microbiome diversity and high-fibre intake are related to lower long-term weight gain. *Int. J. Obes.* . **41**, 1099–1105 (2017).

516. G. Sharon, N. Garg, J. Debelius, R. Knight, P. C. Dorrestein, S. K. Mazmanian, Specialized metabolites from the microbiome in health and disease. *Cell Metab.* **20**, 719–730 (2014).

517. R. Caesar, V. Tremaroli, P. Kovatcheva-Datchary, P. D. Cani, F. Bäckhed, Crosstalk between Gut Microbiota and Dietary Lipids Aggravates WAT Inflammation through TLR Signaling. *Cell Metab.* **22**, 658–668 (2015).

518. S. Ussar, S. Fujisaka, C. R. Kahn, Interactions between host genetics and gut microbiome in diabetes and metabolic syndrome. *Mol Metab.* **5**, 795–803 (2016).

519. C. Ling, L. Groop, Epigenetics: a molecular link between environmental factors and type 2 diabetes. *Diabetes.* **58**, 2718–2725 (2009).

520. P. W. Franks, C. Ling, Epigenetics and obesity: the devil is in the details. *BMC Med.* **8** (2010), p. 88.

521. L. Bouchard, R. Rabasa-Lhoret, M. Faraj, M.-E. Lavoie, J. Mill, L. Pérusse, M.-C. Vohl, Differential epigenomic and transcriptomic responses in subcutaneous adipose tissue between low and high responders to caloric restriction. *Am. J. Clin. Nutr.* **91**, 309–320 (2010).

522. M. A. J. Hullar, B. C. Fu, Diet, the gut microbiome, and epigenetics. *Cancer J.* **20**, 170–175 (2014).

523. R. C. Laker, C. Garde, D. M. Camera, W. J. Smiles, J. R. Zierath, J. A. Hawley, R. Barrès, Transcriptomic and epigenetic responses to short-term nutrient-exercise stress in humans. *Sci. Rep.*

7, 15134 (2017).

524. S. Peleg, C. Feller, A. G. Ladurner, A. Imhof, The Metabolic Impact on Histone Acetylation and Transcription in Ageing. *Trends Biochem. Sci.* **41**, 700–711 (2016).

525. A. E. Ringel, S. A. Tucker, M. C. Haigis, Chemical and Physiological Features of Mitochondrial Acylation. *Mol. Cell.* **72**, 610–624 (2018).

526. K. Aleksandrova, B. Romero-Mosquera, V. Hernandez, Diet, Gut Microbiome and Epigenetics: Emerging Links with Inflammatory Bowel Diseases and Prospects for Management and Prevention. *Nutrients*. **9** (2017), doi:10.3390/nu9090962.

527. K. A. Romano, A. Martinez-Del Campo, K. Kasahara, C. L. Chittim, E. I. Vivas, D. Amador-Noguez, E. P. Balskus, F. E. Rey, Metabolic, Epigenetic, and Transgenerational Effects of Gut Bacterial Choline Consumption. *Cell Host Microbe*. **22**, 279–290.e7 (2017).

528. L. Cai, B. M. Sutter, B. Li, B. P. Tu, Acetyl-CoA induces cell growth and proliferation by promoting the acetylation of histones at growth genes. *Mol. Cell.* **42**, 426–437 (2011).

529. J.-F. Bolduc, L. Hany, C. Barat, M. Ouellet, M. J. Tremblay, Epigenetic Metabolite Acetate Inhibits Class I/II Histone Deacetylases, Promotes Histone Acetylation, and Increases HIV-1 Integration in CD4+ T Cells. *J. Virol.* **91** (2017), doi:10.1128/JVI.01943-16.

530. P. D. Cani, M. Osto, L. Geurts, A. Everard, Involvement of gut microbiota in the development of low-grade inflammation and type 2 diabetes associated with obesity. *Gut Microbes*. **3**, 279–288 (2012).

531. J. J. Seeley, R. G. Baker, G. Mohamed, T. Bruns, M. S. Hayden, S. D. Deshmukh, D. E. Freedberg, S. Ghosh, Induction of innate immune memory via microRNA targeting of chromatin remodelling factors. *Nature*. **559**, 114–119 (2018).

532. Y. Umesaki, Y. Okada, S. Matsumoto, A. Imaoka, H. Setoyama, Segmented filamentous bacteria are indigenous intestinal bacteria that activate intraepithelial lymphocytes and induce MHC class II molecules and fucosyl asialo GM1 glycolipids on the small intestinal epithelial cells in the ex-germ-free mouse. *Microbiol. Immunol.* **39**, 555–562 (1995).

533. J. P.-Y. Ting, J. Trowsdale, Genetic control of MHC class II expression. *Cell.* **109 Suppl**, S21–33 (2002).

534. L. A. Poirier, C. K. Wise, R. R. Delongchamp, R. Sinha, Blood determinations of S-adenosylmethionine, S-adenosylhomocysteine, and homocysteine: correlations with diet. *Cancer Epidemiol. Biomarkers Prev.* **10**, 649–655 (2001).

535. K. A. Krautkramer, F. E. Rey, J. M. Denu, Chemical signaling between gut microbiota and host chromatin: What is your gut really saying? *J. Biol. Chem.* **292**, 8582–8593 (2017).

536. Y.-I. Kim, Nutritional epigenetics: impact of folate deficiency on DNA methylation and colon cancer susceptibility. *J. Nutr.* **135**, 2703–2709 (2005).

537. S. Greenblum, P. J. Turnbaugh, E. Borenstein, Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 594–599 (2012).

538. J. An, A. Rao, M. Ko, TET family dioxygenases and DNA demethylation in stem cells and cancers. *Exp. Mol. Med.* **49**, e323 (2017).

539. D. Bungard, B. J. Fuerth, P.-Y. Zeng, B. Faubert, N. L. Maas, B. Viollet, D. Carling, C. B. Thompson, R. G. Jones, S. L. Berger, Signaling kinase AMPK activates stress-promoted transcription via histone H2B phosphorylation. *Science*. **329**, 1201–1205 (2010).

540. B. Kola, A. B. Grossman, M. Korbonits, The role of AMP-activated protein kinase in obesity. *Front. Horm. Res.* **36**, 198–211 (2008).

541. F. Bäckhed, J. K. Manchester, C. F. Semenkovich, J. I. Gordon, Mechanisms underlying the resistance to diet-induced obesity in germ-free mice. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 979–984 (2007).

542. P. J. Turnbaugh, M. Hamady, T. Yatsunenko, B. L. Cantarel, A. Duncan, R. E. Ley, M. L. Sogin, W. J. Jones, B. A. Roe, J. P. Affourtit, M. Egholm, B. Henrissat, A. C. Heath, R. Knight, J. I. Gordon, A core gut microbiome in obese and lean twins. *Nature*. **457**, 480–484 (2009).

543. J. Gordon, M. Youle, N. Knowlton, F. Rohwer, D. A. Relman, Superorganisms and Holobionts. *Microbe Wash. DC.* **8**, 152–153 (2013).

544. T. J. Carrier, A. M. Reitzel, The Hologenome Across Environments and the Implications of a Host-Associated Microbial Repertoire. *Front. Microbiol.* **8**, 802 (2017).

545. M. Groussin, F. Mazel, J. G. Sanders, C. S. Smillie, S. Lavergne, W. Thuiller, E. J. Alm, Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. *Nat. Commun.* **8**, 14319 (2017).

546. M. van de Guchte, H. M. Blottière, J. Doré, Humans as holobionts: implications for prevention and therapy. *Microbiome*. **6**, 81 (2018).

547. C. A. Hassig, J. K. Tong, S. L. Schreiber, Fiber-derived butyrate and the prevention of colon cancer. *Chem. Biol.* **4**, 783–789 (1997).

548. A. Wahlström, S. I. Sayin, H.-U. Marschall, F. Bäckhed, Intestinal Crosstalk between Bile Acids and Microbiota and Its Impact on Host Metabolism. *Cell Metab.* **24**, 41–50 (2016).

549. B. Paul, S. Barnes, W. Demark-Wahnefried, C. Morrow, C. Salvador, C. Skibola, T. O. Tollefsbol, Influences of diet and the gut microbiome on epigenetic modulation in cancer and other diseases. *Clin. Epigenetics*. **7**, 112 (2015).

550. V. E. Manzo, A. S. Bhatt, The human microbiome in hematopoiesis and hematologic disorders. *Blood*. **126**, 311–318 (2015).

551. B. S. Christmann, T. R. Abrahamsson, C. N. Bernstein, L. W. Duck, P. J. Mannon, G. Berg, B. Björkstén, M. C. Jenmalm, C. O. Elson, Human seroreactivity to gut microbiota antigens. *J. Allergy Clin. Immunol.* **136**, 1378–86.e1–5 (2015).

552. A. Heintz-Buschart, P. May, C. C. Laczny, L. A. Lebrun, C. Bellora, A. Krishna, L. Wampach, J. G. Schneider, A. Hogan, C. de Beaufort, P. Wilmes, Integrated multi-omics of the human gut microbiome in a case study of familial type 1 diabetes. *Nat Microbiol.* **2**, 16180 (2016).

553. E. A. Franzosa, T. Hsu, A. Sirota-Madi, A. Shafquat, G. Abu-Ali, X. C. Morgan, C. Huttenhower,

Sequencing and beyond: integrating molecular "omics" for microbial community profiling. *Nat. Rev. Microbiol.* **13**, 360–372 (2015).

554. I. A. Williamson, J. W. Arnold, L. A. Samsa, L. Gaynor, M. DiSalvo, J. L. Cocchiaro, I. Carroll, M. A. Azcarate-Peril, J. F. Rawls, N. L. Allbritton, S. T. Magness, A High-Throughput Organoid Microinjection Platform to Study Gastrointestinal Microbiota and Luminal Physiology. *Cell Mol Gastroenterol Hepatol.* **6**, 301–319 (2018).

555. E. E. Nelson, A. E. Guyer, The development of the ventral prefrontal cortex and social flexibility. *Dev. Cogn. Neurosci.* **1**, 233–245 (2011).

556. E. van Nood, A. Vrieze, M. Nieuwdorp, S. Fuentes, E. G. Zoetendal, W. M. de Vos, C. E. Visser, E. J. Kuijper, J. F. W. M. Bartelsman, J. G. P. Tijssen, P. Speelman, M. G. W. Dijkgraaf, J. J. Keller, Duodenal infusion of donor feces for recurrent Clostridium difficile. *N. Engl. J. Med.* **368**, 407–415 (2013).

557. R. S. Kootte, E. Levin, J. Salojärvi, L. P. Smits, A. V. Hartstra, S. D. Udayappan, G. Hermes, K. E. Bouter, A. M. Koopen, J. J. Holst, F. K. Knop, E. E. Blaak, J. Zhao, H. Smidt, A. C. Harms, T. Hankemeijer, J. J. G. H. M. Bergman, H. A. Romijn, F. G. Schaap, S. W. M. Olde Damink, M. T. Ackermans, G. M. Dallinga-Thie, E. Zoetendal, W. M. de Vos, M. J. Serlie, E. S. G. Stroes, A. K. Groen, M. Nieuwdorp, Improvement of Insulin Sensitivity after Lean Donor Feces in Metabolic Syndrome Is Driven by Baseline Intestinal Microbiota Composition. *Cell Metab.* **26**, 611–619.e6 (2017).

558. S. Gupta, E. Allen-Vercoe, E. O. Petrof, Fecal microbiota transplantation: in perspective. *Therap. Adv. Gastroenterol.* **9**, 229–239 (2016).

559. D.-W. Kang, J. B. Adams, A. C. Gregory, T. Borody, L. Chittick, A. Fasano, A. Khoruts, E. Geis, J. Maldonado, S. McDonough-Means, E. L. Pollard, S. Roux, M. J. Sadowsky, K. S. Lipson, M. B. Sullivan, J. G. Caporaso, R. Krajmalnik-Brown, Microbiota Transfer Therapy alters gut ecosystem and improves gastrointestinal and autism symptoms: an open-label study. *Microbiome.* **5**, 10 (2017).

560. E. G. Pamer, Fecal microbiota transplantation: effectiveness, complexities, and lingering concerns. *Mucosal Immunol.* **7**, 210–214 (2014).

561. N. Koppel, V. Maini Rekdal, E. P. Balskus, Chemical transformation of xenobiotics by the human gut microbiota. *Science.* **356** (2017), doi:10.1126/science.aag2770.

562. M. I. Bhat, R. Kapila, Dietary metabolites derived from gut microbiota: critical modulators of epigenetic changes in mammals. *Nutr. Rev.* **75**, 374–389 (2017).

563. B. P. Landry, J. J. Tabor, Engineering Diagnostic and Therapeutic Gut Bacteria. *Microbiol Spectr.* **5** (2017), doi:10.1128/microbiolspec.BAD-0020-2017.

564. P. Manrique, M. Dills, M. J. Young, The Human Gut Phage Community and Its Implications for Health and Disease. *Viruses.* **9** (2017), doi:10.3390/v9060141.

565. L. Maier, M. Pruteanu, M. Kuhn, G. Zeller, A. Telzerow, E. E. Anderson, A. R. Brochado, K. C. Fernandez, H. Dose, H. Mori, K. R. Patil, P. Bork, A. Typas, Extensive impact of non-antibiotic drugs on human gut bacteria. *Nature.* **555**, 623–628 (2018).

566. R. Nagpal, S. Wang, S. Ahmadi, J. Hayes, J. Gagliano, S. Subashchandrabose, D. W. Kitzman, T. Becton, R. Read, H. Yadav, Human-origin probiotic cocktail increases short-chain fatty acid

production via modulation of mice and human gut microbiome. *Sci. Rep.* **8**, 12649 (2018).

567. K. Tillisch, J. Labus, L. Kilpatrick, Z. Jiang, J. Stains, B. Ebrat, D. Guyonnet, S. Legrain-Raspaud, B. Trotin, B. Naliboff, E. A. Mayer, Consumption of fermented milk product with probiotic modulates brain activity. *Gastroenterology*. **144**, 1394–401, 1401.e1–4 (2013).

568. M. Pane, A. Amoruso, F. Deidda, T. Graziano, S. Allesina, L. Mogna, Gut Microbiota, Probiotics, and Sport: From Clinical Evidence to Agonistic Performance. *J. Clin. Gastroenterol.* **52 Suppl 1, Proceedings from the 9th Probiotics, Prebiotics and New Foods, Nutraceuticals and Botanicals for Nutrition & Human and Microbiota Health Meeting, held in Rome, Italy from September 10 to 12, 2017**, S46–S49 (2018).

569. G. Reid, J. Burton, Use of Lactobacillus to prevent infection by pathogenic bacteria. *Microbes Infect.* **4**, 319–324 (2002).

570. A. Langdon, N. Crook, G. Dantas, The effects of antibiotics on the microbiome throughout development and alternative approaches for therapeutic modulation. *Genome Med.* **8**, 39 (2016).

571. L. Zhao, F. Zhang, X. Ding, G. Wu, Y. Y. Lam, X. Wang, H. Fu, X. Xue, C. Lu, J. Ma, L. Yu, C. Xu, Z. Ren, Y. Xu, S. Xu, H. Shen, X. Zhu, Y. Shi, Q. Shen, W. Dong, R. Liu, Y. Ling, Y. Zeng, X. Wang, Q. Zhang, J. Wang, L. Wang, Y. Wu, B. Zeng, H. Wei, M. Zhang, Y. Peng, C. Zhang, Gut bacteria selectively promoted by dietary fibers alleviate type 2 diabetes. *Science*. **359**, 1151–1156 (2018).

572. Z. Dai, V. Ramesh, J. W. Locasale, The evolving metabolic landscape of chromatin biology and epigenetics. *Nat. Rev. Genet.* **21**, 737–753 (2020).

573. S. A. Quinodoz, N. Ollikainen, B. Tabak, A. Palla, J. M. Schmidt, E. Detmar, M. M. Lai, A. A. Shishkin, P. Bhat, Y. Takei, V. Trinh, E. Aznauryan, P. Russell, C. Cheng, M. Jovanovic, A. Chow, L. Cai, P. McDonel, M. Garber, M. Guttman, Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell.* **174**, 744–757.e24 (2018).

574. F. Ciabrelli, G. Cavalli, Chromatin-driven behavior of topologically associating domains. *J. Mol. Biol.* **427**, 608–625 (2015).

575. S. Maegawa, Y. Lu, T. Tahara, J. T. Lee, J. Madzo, S. Liang, J. Jelinek, R. J. Colman, J.-P. J. Issa, Caloric restriction delays age-related methylation drift. *Nat. Commun.* **8**, 539 (2017).

576. J. Fan, K. A. Krautkramer, J. L. Feldman, J. M. Denu, Metabolic regulation of histone post-translational modifications. *ACS Chem. Biol.* **10**, 95–108 (2015).

577. B. Cascella, L. M. Mirica, Kinetic analysis of iron-dependent histone demethylases: α-ketoglutarate substrate inhibition and potential relevance to the regulation of histone demethylation in cancer cells. *Biochemistry*. **51**, 8699–8701 (2012).

578. M. Xiao, H. Yang, W. Xu, S. Ma, H. Lin, H. Zhu, L. Liu, Y. Liu, C. Yang, Y. Xu, S. Zhao, D. Ye, Y. Xiong, K.-L. Guan, Inhibition of α-KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors. *Genes Dev.* **26**, 1326–1338 (2012).

579. G. S. Ducker, J. D. Rabinowitz, One-Carbon Metabolism in Health and Disease. *Cell Metab.* **25**, 27–42 (2017).

580. S.-I. Imai, L. Guarente, NAD+ and sirtuins in aging and disease. *Trends Cell Biol.* **24**, 464–471 (2014).

581. P. Mews, G. Donahue, A. M. Drake, V. Luczak, T. Abel, S. L. Berger, Acetyl-CoA synthetase regulates histone acetylation and hippocampal memory. *Nature*. **546**, 381–386 (2017).

582. K. S. Crider, T. P. Yang, R. J. Berry, L. B. Bailey, Folate and DNA methylation: a review of molecular mechanisms and the evidence for folate's role. *Adv. Nutr.* **3**, 21–38 (2012).

583. L. Chen, G. S. Ducker, W. Lu, X. Teng, J. D. Rabinowitz, An LC-MS chemical derivatization method for the measurement of five different one-carbon states of cellular tetrahydrofolate. *Anal. Bioanal. Chem.* **409**, 5955–5964 (2017).

584. L. D'Ulivo, L. Yang, J. Ding, E. Pagliano, D. M. Leek, M.-P. Thibeault, Z. Mester, Determination of cyanocobalamin by isotope dilution LC-MS/MS. *Anal. Chim. Acta.* **990**, 103–109 (2017).

585. S. H. Kirsch, J.-P. Knapp, J. Geisel, W. Herrmann, R. Obeid, Simultaneous quantification of S-adenosyl methionine and S-adenosyl homocysteine in human plasma by stable-isotope dilution ultra performance liquid chromatography tandem mass spectrometry. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **877**, 3865–3870 (2009).

586. X. Liu, S. Sadhukhan, S. Sun, G. R. Wagner, M. D. Hirschey, L. Qi, H. Lin, J. W. Locasale, High-Resolution Metabolomics with Acyl-CoA Profiling Reveals Widespread Remodeling in Response to Diet. *Mol. Cell. Proteomics*. **14**, 1489–1500 (2015).

587. R. H. G. Wright, A. Lioutas, F. Le Dily, D. Soronellas, A. Pohl, J. Bonet, A. S. Nacht, S. Samino, J. Font-Mateu, G. P. Vicent, M. Wierer, M. A. Trabado, C. Schelhorn, C. Carolis, M. J. Macias, O. Yanes, B. Oliva, M. Beato, ADP-ribose-derived nuclear ATP synthesis by NUDIX5 is required for chromatin remodeling. *Science*. **352**, 1221–1225 (2016).

588. M. Schittmayer, R. Birner-Gruenberger, N. Zamboni, Quantification of Cellular Folate Species by LC-MS after Stabilization by Derivatization. *Anal. Chem.* **90**, 7349–7356 (2018).

589. C. Lotti, J. Rubert, F. Fava, K. Tuohy, F. Mattivi, U. Vrhovsek, Development of a fast and cost-effective gas chromatography-mass spectrometry method for the quantification of short-chain and medium-chain fatty acids in human biofluids. *Anal. Bioanal. Chem.* **409**, 5555–5567 (2017).

590. M. Yang, P. J. Pollard, Succinate: a new epigenetic hacker. *Cancer Cell.* **23** (2013), pp. 709–711.

591. O. Shuvalov, A. Petukhov, A. Daks, O. Fedorova, E. Vasileva, N. A. Barlev, One-carbon metabolism and nucleotide biosynthesis as attractive targets for anticancer therapy. *Oncotarget*. **8**, 23955–23977 (2017).

592. E. C. Rush, P. Katre, C. S. Yajnik, Vitamin B12: one carbon metabolism, fetal growth and programming for chronic disease. *Eur. J. Clin. Nutr.* **68**, 2–7 (2014).

593. O. D. K. Maddocks, C. F. Labuschagne, P. D. Adams, K. H. Vousden, Serine Metabolism Supports the Methionine Cycle and DNA/RNA Methylation through De Novo ATP Synthesis in Cancer Cells. *Mol. Cell.* **61**, 210–221 (2016).

594. M. Zeng, H. Cao, Fast quantification of short chain fatty acids and ketone bodies by liquid chromatography-tandem mass spectrometry after facile derivatization coupled with liquid-liquid

extraction. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **1083**, 137–145 (2018).

595. S. G. Lamarre, L. MacMillan, G. P. Morrow, E. Randell, T. Pongnopparat, M. E. Brosnan, J. T. Brosnan, An isotope-dilution, GC-MS assay for formate and its application to human and animal metabolism. *Amino Acids*. **46**, 1885–1891 (2014).

596. J. H. Cummings, E. W. Pomare, W. J. Branch, C. P. Naylor, G. T. Macfarlane, Short chain fatty acids in human large intestine, portal, hepatic and venous blood. *Gut*. **28**, 1221–1227 (1987).

597. E. R. Hughes, M. G. Winter, B. A. Duerkop, L. Spiga, T. Furtado de Carvalho, W. Zhu, C. C. Gillis, L. Büttner, M. P. Smoot, C. L. Behrendt, S. Cherry, R. L. Santos, L. V. Hooper, S. E. Winter, Microbial Respiration and Formate Oxidation as Metabolic Signatures of Inflammation-Associated Dysbiosis. *Cell Host Microbe*. **21**, 208–219 (2017).

598. M. Pietzke, J. Meiser, A. Vazquez, Formate metabolism in health and disease. *Mol Metab*. **33**, 23–37 (2020).

599. J. Boyle, Lehninger principles of biochemistry (4th ed.): Nelson, D., and Cox, M. *Biochemistry and Molecular Biology Education*. **33** (2005), pp. 74–75.

600. T. Mashima, H. Seimiya, T. Tsuruo, De novo fatty-acid synthesis and related pathways as molecular targets for cancer therapy. *Br. J. Cancer*. **100**, 1369–1372 (2009).

601. W. Yu, Z. Wang, K. Zhang, Z. Chi, T. Xu, D. Jiang, S. Chen, W. Li, X. Yang, X. Zhang, Y. Wu, D. Wang, One-Carbon Metabolism Supports S-Adenosylmethionine and Histone Methylation to Drive Inflammatory Macrophages. *Mol. Cell*. **75**, 1147–1160.e5 (2019).

602. K. A. Krautkramer, J. Fan, F. Bäckhed, Gut microbial metabolites as multi-kingdom intermediates. *Nat. Rev. Microbiol.* **19**, 77–94 (2021).

603. H. N. Sanchez, J. B. Moroney, H. Gan, T. Shen, J. L. Im, T. Li, J. R. Taylor, H. Zan, P. Casali, B cell-intrinsic epigenetic modulation of antibody responses by dietary fiber-derived short-chain fatty acids. *Nat. Commun.* **11**, 60 (2020).

604. R. Berni Canani, M. Di Costanzo, L. Leone, The epigenetic effects of butyrate: potential therapeutic implications for clinical practice. *Clin. Epigenetics*. **4**, 4 (2012).

605. X. Gao, S.-H. Lin, F. Ren, J.-T. Li, J.-J. Chen, C.-B. Yao, H.-B. Yang, S.-X. Jiang, G.-Q. Yan, D. Wang, Y. Wang, Y. Liu, Z. Cai, Y.-Y. Xu, J. Chen, W. Yu, P.-Y. Yang, Q.-Y. Lei, Acetate functions as an epigenetic metabolite to promote lipid synthesis under hypoxia. *Nat. Commun.* **7**, 11960 (2016).

606. I. R. Miousse, R. Pathak, S. Garg, C. M. Skinner, S. Melnyk, O. Pavliv, H. Hendrickson, R. D. Landes, A. Lumen, A. J. Tackett, N. E. P. Deutz, M. Hauer-Jensen, I. Koturbash, Short-term dietary methionine supplementation affects one-carbon metabolism and DNA methylation in the mouse gut and leads to altered microbiome profiles, barrier function, gene expression and histomorphology. *Genes Nutr.* **12**, 22 (2017).

607. V. T. Pham, S. Fehlbaum, N. Seifert, N. Richard, M. J. Bruins, W. Sybesma, A. Rehman, R. E. Steinert, Effects of colon-targeted vitamins on the composition and metabolic activity of the human gut microbiome- a pilot study. *Gut Microbes*. **13**, 1–20 (2021).

608. I. Huber-Ruano, E. Calvo, J. Mayneris-Perxachs, M.-M. Rodríguez-Peña, V. Ceperuelo-Mallafré, L. Cedó, C. Núñez-Roa, J. Miro-Blanch, M. Arnoriaga-Rodríguez, A. Balvay, C. Maudet, P.

García-Roves, O. Yanes, S. Rabot, G. M. Grimaud, A. De Prisco, A. Amoruso, J. M. Fernández-Real, J. Vendrell, S. Fernández-Veledo, Orally administered Odoribacter laneus improves glucose control and inflammatory profile in obese mice by depleting circulating succinate. *Microbiome*. **10**, 135 (2022).

609. M. A. Lauterbach, J. E. Hanke, M. Serefidou, M. S. J. Mangan, C.-C. Kolbe, T. Hess, M. Rothe, R. Kaiser, F. Hoss, J. Gehlen, G. Engels, M. Kreutzenbeck, S. V. Schmidt, A. Christ, A. Imhof, K. Hiller, E. Latz, Toll-like Receptor Signaling Rewires Macrophage Metabolism and Promotes Histone Acetylation via ATP-Citrate Lyase. *Immunity*. **51**, 997–1011.e7 (2019).

610. C. A. Butts, G. Paturi, M. H. Tavendale, D. Hedderley, H. M. Stoklosinski, T. D. Herath, D. Rosendale, N. C. Roy, J. A. Monro, J. Ansell, The fate of (13)C-labelled and non-labelled inulin predisposed to large bowel fermentation in rats. *Food Funct.* **7**, 1825–1832 (2016).

611. P. J. Lund, L. A. Gates, M. Leboeuf, S. A. Smith, L. Chau, M. Lopes, E. S. Friedman, Y. Saiman, M. S. Kim, C. A. Shoffler, C. Petucci, C. D. Allis, G. D. Wu, B. A. Garcia, Stable isotope tracing in vivo reveals a metabolic bridge linking the microbiota to host histone acetylation. *Cell Rep.* **41**, 111809 (2022).

612. M. Yuan, S. B. Breitkopf, X. Yang, J. M. Asara, A positive/negative ion-switching, targeted mass spectrometry-based metabolomics platform for bodily fluids, cells, and fresh and fixed tissue. *Nat. Protoc.* **7**, 872–881 (2012).

613. M. Oliva, M. Muñoz-Aguirre, S. Kim-Hellmuth, V. Wucher, A. D. H. Gewirtz, D. J. Cotter, P. Parsana, S. Kasela, B. Balliu, A. Viñuela, S. E. Castel, P. Mohammadi, F. Aguet, Y. Zou, E. A. Khramtsova, A. D. Skol, D. Garrido-Martín, F. Reverter, A. Brown, P. Evans, E. R. Gamazon, A. Payne, R. Bonazzola, A. N. Barbeira, A. R. Hamel, A. Martinez-Perez, J. M. Soria, GTEx Consortium, B. L. Pierce, M. Stephens, E. Eskin, E. T. Dermitzakis, A. V. Segrè, H. K. Im, B. E. Engelhardt, K. G. Ardlie, S. B. Montgomery, A. J. Battle, T. Lappalainen, R. Guigó, B. E. Stranger, The impact of sex on gene expression across human tissues. *Science*. **369** (2020), doi:10.1126/science.aba3066.

614. E. Hall, P. Volkov, T. Dayeh, J. L. S. Esguerra, S. Salö, L. Eliasson, T. Rönn, K. Bacos, C. Ling, Sex differences in the genome-wide DNA methylation pattern and impact on gene expression, microRNA levels and insulin secretion in human pancreatic islets. *Genome Biol.* **15**, 522 (2014).

615. H. MacKay, C. J. Gunasekara, K.-Y. Yam, D. Srisai, H. K. Yalamanchili, Y. Li, R. Chen, C. Coarfa, R. A. Waterland, Sex-specific epigenetic development in the mouse hypothalamic arcuate nucleus pinpoints human genomic regions associated with body mass index. *Sci. Adv.* **8** (2022), doi:10.1126/sciadv.abo3991.

616. A. Gao, J. Su, R. Liu, S. Zhao, W. Li, X. Xu, D. Li, J. Shi, B. Gu, J. Zhang, Q. Li, X. Wang, Y. Zhang, Y. Xu, J. Lu, G. Ning, J. Hong, Y. Bi, W. Gu, J. Wang, W. Wang, Sexual dimorphism in glucose metabolism is shaped by androgen-driven gut microbiome. *Nat. Commun.* **12**, 7080 (2021).

617. E. C. Jenkins, N. Shah, M. Gomez, G. Casalena, D. Zhao, T. C. Kenny, S. R. Guariglia, G. Manfredi, D. Germain, Proteasome mapping reveals sexual dimorphism in tissue-specific sensitivity to protein aggregations. *EMBO Rep.* **21**, e48978 (2020).

618. X. Yang, E. E. Schadt, S. Wang, H. Wang, A. P. Arnold, L. Ingram-Drake, T. A. Drake, A. J. Lusis, Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome Res.* **16**, 995–1004 (2006).

619. D. Zheng, J. Trynda, C. Williams, J. A. Vold, J. H. Nguyen, D. M. Harnois, S. P. Bagaria, S. A. McLaughlin, Z. Li, Sexual dimorphism in the incidence of human cancers. *BMC Cancer*. **19**, 684 (2019).

620. A. G. Huebschmann, R. R. Huxley, W. M. Kohrt, P. Zeitler, J. G. Regensteiner, J. E. B. Reusch, Sex differences in the burden of type 2 diabetes and cardiovascular risk across the life course. *Diabetologia*. **62**, 1761–1772 (2019).

621. F. Garawi, K. Devries, N. Thorogood, R. Uauy, Global differences between women and men in the prevalence of obesity: is there an association with gender inequality? *Eur. J. Clin. Nutr.* **68**, 1101–1106 (2014).

622. M. Severs, L. M. Spekhorst, M.-J. J. Mangen, G. Dijkstra, M. Löwenberg, F. Hoentjen, A. E. van der Meulen-de Jong, M. Pierik, C. Y. Ponsioen, G. Bouma, J. C. van der Woude, M. E. van der Valk, M. J. L. Romberg-Camps, C. H. M. Clemens, P. van de Meeberg, N. Mahmmod, J. Jansen, B. Jharap, R. K. Weersma, B. Oldenburg, E. A. M. Festen, H. H. Fidder, Sex-Related Differences in Patients With Inflammatory Bowel Disease: Results of 2 Prospective Cohort Studies. *Inflamm. Bowel Dis.* **24**, 1298–1306 (2018).

623. L. L. Barnes, R. S. Wilson, J. L. Bienias, J. A. Schneider, D. A. Evans, D. A. Bennett, Sex differences in the clinical manifestations of Alzheimer disease pathology. *Arch. Gen. Psychiatry*. **62**, 685–691 (2005).

624. S. T. Ngo, F. J. Steyn, P. A. McCombe, Gender differences in autoimmune disease. *Front. Neuroendocrinol.* **35**, 347–369 (2014).

625. Website, (available at https://data.europa.eu/doi/10.2777/316197).

626. NIH Policy and Guidelines on The Inclusion of Women and Minorities as Subjects in Clinical Research, (available at https://grants.nih.gov/policy/inclusion/women-and-minorities/guidelines.htm).

627. J. A. Clayton, F. S. Collins, Policy: NIH to balance sex in cell and animal studies. *Nature*. **509** (2014), pp. 282–283.

628. R. M. Shansky, Are hormones a "female problem" for animal research? *Science*. **364**, 825–826 (2019).

629. L. R. Miller, C. Marks, J. B. Becker, P. D. Hurn, W.-J. Chen, T. Woodruff, M. M. McCarthy, F. Sohrabji, L. Schiebinger, C. L. Wetherington, S. Makris, A. P. Arnold, G. Einstein, V. M. Miller, K. Sandberg, S. Maier, T. L. Cornelison, J. A. Clayton, Considering sex as a biological variable in preclinical research. *FASEB J.* **31**, 29–34 (2017).

630. J. A. Clayton, Applying the new SABV (sex as a biological variable) policy to research and clinical care. *Physiol. Behav.* **187**, 2–5 (2018).

631. M. E. Arnegard, L. A. Whitten, C. Hunter, J. A. Clayton, Sex as a Biological Variable: A 5-Year Progress Report and Call to Action. *J. Womens. Health* . **29**, 858–864 (2020).

632. J. B. Becker, B. J. Prendergast, J. W. Liang, Female rats are not more variable than male rats: a meta-analysis of neuroscience studies. *Biol. Sex Differ.* **7**, 34 (2016).

633. F. Mauvais-Jarvis, A. P. Arnold, K. Reue, A Guide for the Design of Pre-clinical Studies on Sex

Differences in Metabolism. *Cell Metab.* **25**, 1216–1230 (2017).

634. E. Brady, M. W. Nielsen, J. P. Andersen, S. Oertelt-Prigione, Lack of consideration of sex and gender in COVID-19 clinical studies. *Nat. Commun.* **12**, 1–6 (2021).

635. M. Wlodarska, A. D. Kostic, R. J. Xavier, An integrative view of microbiome-host interactions in inflammatory bowel diseases. *Cell Host Microbe.* **17**, 577–591 (2015).

636. I. Cho, M. J. Blaser, The human microbiome: at the interface of health and disease. *Nat. Rev. Genet.* **13**, 260–270 (2012).

637. S. C. Shin, S.-H. Kim, H. You, B. Kim, A. C. Kim, K.-A. Lee, J.-H. Yoon, J.-H. Ryu, W.-J. Lee, Drosophila microbiome modulates host developmental and metabolic homeostasis via insulin signaling. *Science.* **334**, 670–674 (2011).

638. C. Heintz, W. Mair, You are what you host: microbiome modulation of the aging process. *Cell.* **156**, 408–411 (2014).

639. F. Sommer, F. Bäckhed, The gut microbiota--masters of host development and physiology. *Nat. Rev. Microbiol.* **11**, 227–238 (2013).

640. M. Bredon, J. Dittmer, C. Noël, B. Moumen, D. Bouchon, Lignocellulose degradation at the holobiont level: teamwork in a keystone soil invertebrate. *Microbiome.* **6**, 162 (2018).

641. S. M. Murga-Garrido, Q. Hong, T.-W. L. Cross, E. R. Hutchison, J. Han, S. P. Thomas, E. I. Vivas, J. Denu, D. G. Ceschin, Z.-Z. Tang, F. E. Rey, Gut microbiome variation modulates the effects of dietary fiber on host metabolism. *Microbiome.* **9**, 117 (2021).

642. P. J. Turnbaugh, R. E. Ley, M. A. Mahowald, V. Magrini, E. R. Mardis, J. I. Gordon, An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature.* **444**, 1027–1031 (2006).

643. J. L. Gehrig, S. Venkatesh, H.-W. Chang, M. C. Hibberd, V. L. Kung, J. Cheng, R. Y. Chen, S. Subramanian, C. A. Cowardin, M. F. Meier, D. O'Donnell, M. Talcott, L. D. Spears, C. F. Semenkovich, B. Henrissat, R. J. Giannone, R. L. Hettich, O. Ilkayeva, M. Muehlbauer, C. B. Newgard, C. Sawyer, R. D. Head, D. A. Rodionov, A. A. Arzamasov, S. A. Leyn, A. L. Osterman, M. I. Hossain, M. Islam, N. Choudhury, S. A. Sarker, S. Huq, I. Mahmud, I. Mostafa, M. Mahfuz, M. J. Barratt, T. Ahmed, J. I. Gordon, Effects of microbiota-directed foods in gnotobiotic animals and undernourished children. *Science.* **365** (2019), doi:10.1126/science.aau4732.

644. R. C. Bortolin, A. R. Vargas, J. Gasparotto, P. R. Chaves, C. E. Schnorr, K. B. Martinello, A. K. Silveira, T. K. Rabelo, D. P. Gelain, J. C. F. Moreira, A new animal diet based on human Western diet is a robust diet-induced obesity model: comparison to high-fat and cafeteria diets in term of metabolic and gut microbiota disruption. *Int. J. Obes.* . **42**, 525–534 (2018).

645. S. Gurwara, N. J. Ajami, A. Jang, F. C. Hessel, L. Chen, S. Plew, Z. Wang, D. Y. Graham, C. Hair, D. L. White, J. Kramer, T. Kourkoumpetis, K. Hoffman, R. Cole, J. Hou, N. Husain, M. Jarbrink-Sehgal, R. Hernaez, F. Kanwal, G. Ketwaroo, R. Shah, M. Velez, Y. Natarajan, H. B. El-Serag, J. F. Petrosino, L. Jiao, Dietary Nutrients Involved in One-Carbon Metabolism and Colonic Mucosa-Associated Gut Microbiome in Individuals with an Endoscopically Normal Colon. *Nutrients.* **11** (2019), doi:10.3390/nu11030613.

646. L. Yurkovetskiy, M. Burrows, A. A. Khan, L. Graham, P. Volchkov, L. Becker, D. Antonopoulos, Y.

Umesaki, A. V. Chervonsky, Gender bias in autoimmunity is influenced by microbiota. *Immunity*. **39**, 400–412 (2013).

647. E. Org, M. Mehrabian, B. W. Parks, P. Shipkova, X. Liu, T. A. Drake, A. J. Lusis, Sex differences and hormonal effects on gut microbiota composition in mice. *Gut Microbes*. **7**, 313–322 (2016).

648. J. Mayneris-Perxachs, M. Arnoriaga-Rodríguez, D. Luque-Córdoba, F. Priego-Capote, V. Pérez-Brocal, A. Moya, A. Burokas, R. Maldonado, J.-M. Fernández-Real, Gut microbiota steroid sexual dimorphism and its impact on gonadal steroids: influences of obesity and menopausal status. *Microbiome*. **8**, 136 (2020).

649. T.-W. L. Cross, K. Kasahara, F. E. Rey, Sexual dimorphism of cardiometabolic dysfunction: Gut microbiome in the play? *Mol Metab*. **15**, 70–81 (2018).

650. F. Fransen, A. A. van Beek, T. Borghuis, B. Meijer, F. Hugenholtz, C. van der Gaast-de Jongh, H. F. Savelkoul, M. I. de Jonge, M. M. Faas, M. V. Boekschoten, H. Smidt, S. El Aidy, P. de Vos, The Impact of Gut Microbiota on Gender-Specific Differences in Immunity. *Front. Immunol.* **8**, 754 (2017).

651. J. M. Edwards, S. Roy, J. C. Tomcho, Z. J. Schreckenberger, S. Chakraborty, N. R. Bearss, P. Saha, C. G. McCarthy, M. Vijay-Kumar, B. Joe, C. F. Wenceslau, Microbiota are critical for vascular physiology: Germ-free status weakens contractility and induces sex-specific vascular remodeling in mice. *Vascul. Pharmacol.* **125-126**, 106633 (2020).

652. C. M. Lopes-Ramos, C.-Y. Chen, M. L. Kuijjer, J. N. Paulson, A. R. Sonawane, M. Fagny, J. Platig, K. Glass, J. Quackenbush, D. L. DeMeo, Sex Differences in Gene Expression and Regulatory Networks across 29 Human Tissues. *Cell Rep.* **31**, 107795 (2020).

653. R. J. G. Hartman, D. M. C. Kapteijn, S. Haitjema, M. N. Bekker, M. Mokry, G. Pasterkamp, M. Civelek, H. M. den Ruijter, Intrinsic transcriptomic sex differences in human endothelial cells at birth and in adults are associated with coronary artery disease targets. *Sci. Rep.* **10**, 12367 (2020).

654. D. Guneykaya, A. Ivanov, D. P. Hernandez, V. Haage, B. Wojtas, N. Meyer, M. Maricos, P. Jordan, A. Buonfiglioli, B. Gielniewski, N. Ochocka, C. Cömert, C. Friedrich, L. S. Artiles, B. Kaminska, P. Mertins, D. Beule, H. Kettenmann, S. A. Wolf, Transcriptional and Translational Differences of Microglia from Male and Female Brains. *Cell Rep.* **24**, 2773–2783.e6 (2018).

655. M. Melé, P. G. Ferreira, F. Reverter, D. S. DeLuca, J. Monlong, M. Sammeth, T. R. Young, J. M. Goldmann, D. D. Pervouchine, T. J. Sullivan, R. Johnson, A. V. Segrè, S. Djebali, A. Niarchou, GTEx Consortium, F. A. Wright, T. Lappalainen, M. Calvo, G. Getz, E. T. Dermitzakis, K. G. Ardlie, R. Guigó, Human genomics. The human transcriptome across tissues and individuals. *Science*. **348**, 660–665 (2015).

656. D. Trabzuni, A. Ramasamy, S. Imran, R. Walker, C. Smith, M. E. Weale, J. Hardy, M. Ryten, North American Brain Expression Consortium, Widespread sex differences in gene expression and splicing in the adult human brain. *Nat. Commun.* **4**, 2771 (2013).

657. D. Chicco, G. Agapito, Nine quick tips for pathway enrichment analysis. *PLoS Comput. Biol.* **18**, e1010348 (2022).

658. V. V. Flis, G. Daum, Lipid transport between the endoplasmic reticulum and mitochondria. *Cold Spring Harb. Perspect. Biol.* **5** (2013), doi:10.1101/cshperspect.a013235.

659. W. L. Miller, R. J. Auchus, The molecular biology, biochemistry, and physiology of human steroidogenesis and its disorders. *Endocr. Rev.* **32**, 81–151 (2011).

660. L. Schiffer, L. Barnard, E. S. Baranowski, L. C. Gilligan, A. E. Taylor, W. Arlt, C. H. L. Shackleton, K.-H. Storbeck, Human steroid biosynthesis, metabolism and excretion are differentially reflected by serum and urine steroid metabolomes: A comprehensive review. *J. Steroid Biochem. Mol. Biol.* **194**, 105439 (2019).

661. A. D. Read, R. E. Bentley, S. L. Archer, K. J. Dunham-Snary, Mitochondrial iron-sulfur clusters: Structure, function, and an emerging role in vascular biology. *Redox Biol.* **47**, 102164 (2021).

662. A. Razin, H. Cedar, DNA methylation and gene expression. *Microbiol. Rev.* **55**, 451–458 (1991).

663. M. Kulis, M. Esteller, DNA methylation and cancer. *Adv. Genet.* **70**, 27–56 (2010).

664. J. R. Wagner, S. Busche, B. Ge, T. Kwan, T. Pastinen, M. Blanchette, The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. *Genome Biol.* **15**, R37 (2014).

665. I. Rauluseviciute, F. Drabløs, M. B. Rye, DNA hypermethylation associated with upregulated gene expression in prostate cancer demonstrates the diversity of epigenetic regulation. *BMC Med. Genomics*. **13**, 6 (2020).

666. J. Smith, S. Sen, R. J. Weeks, M. R. Eccles, A. Chatterjee, Promoter DNA Hypermethylation and Paradoxical Gene Activation. *Trends Cancer Res.* **6**, 392–406 (2020).

667. G. Han, H. Luong, S. Vaishnava, Low abundance members of the gut microbiome exhibit high immunogenicity. *Gut Microbes*. **14**, 2104086 (2022).

668. J. Benjamino, S. Lincoln, R. Srivastava, J. Graf, Low-abundant bacteria drive compositional changes in the gut microbiota after dietary alteration. *Microbiome*. **6**, 86 (2018).

669. S. C. Di Rienzi, I. Sharon, K. C. Wrighton, O. Koren, L. A. Hug, B. C. Thomas, J. K. Goodrich, J. T. Bell, T. D. Spector, J. F. Banfield, R. E. Ley, The human gut and groundwater harbor non-photosynthetic bacteria belonging to a new candidate phylum sibling to Cyanobacteria. *Elife*. **2**, e01102 (2013).

670. M. J. Blaser, S. Falkow, What are the consequences of the disappearing human microbiota? *Nat. Rev. Microbiol.* **7**, 887–894 (2009).

671. R. B. Canani, M. D. Costanzo, L. Leone, M. Pedata, R. Meli, A. Calignano, Potential beneficial effects of butyrate in intestinal and extraintestinal diseases. *World J. Gastroenterol.* **17**, 1519–1528 (2011).

672. S. Deleu, K. Machiels, J. Raes, K. Verbeke, S. Vermeire, Short chain fatty acids and its producing organisms: An overlooked therapy for IBD? *EBioMedicine*. **66**, 103293 (2021).

673. X. Gu, J. X. Y. Sim, W. L. Lee, L. Cui, Y. F. Z. Chan, E. D. Chang, Y. E. Teh, A.-N. Zhang, F. Armas, F. Chandra, H. Chen, S. Zhao, Z. Lee, J. R. Thompson, E. E. Ooi, J. G. Low, E. J. Alm, S. Kalimuddin, Gut Ruminococcaceae levels at baseline correlate with risk of antibiotic-associated diarrhea. *iScience*. **25**, 103644 (2022).

674. S. Rabot, M. Membrez, A. Bruneau, P. Gérard, T. Harach, M. Moser, F. Raymond, R. Mansourian,

C. J. Chou, Germ-free C57BL/6J mice are resistant to high-fat-diet-induced insulin resistance and have altered cholesterol metabolism. *FASEB J.* **24**, 4948–4959 (2010).

675. U. S. Pettersson, T. B. Waldén, P.-O. Carlsson, L. Jansson, M. Phillipson, Female mice are protected against high-fat diet induced metabolic syndrome and increase the regulatory T cell population in adipose tissue. *PLoS One*. **7**, e46057 (2012).

676. W. Yu, Q. Lei, L. Yang, G. Qin, S. Liu, D. Wang, Y. Ping, Y. Zhang, Contradictory roles of lipid metabolism in immune response within the tumor microenvironment. *J. Hematol. Oncol.* **14**, 187 (2021).

677. E. K. F. Donahue, E. M. Ruark, K. Burkewitz, Fundamental roles for inter-organelle communication in aging. *Biochem. Soc. Trans.* **50**, 1389–1402 (2022).

678. M. Petkovic, C. E. O'Brien, Y. N. Jan, Interorganelle communication, aging, and neurodegeneration. *Genes Dev.* **35**, 449–469 (2021).

679. K. Todkar, L. Chikhi, M. Germain, Mitochondrial interaction with the endosomal compartment in endocytosis and mitochondrial transfer. *Mitochondrion*. **49**, 284–288 (2019).

680. A. Angajala, S. Lim, J. B. Phillips, J.-H. Kim, C. Yates, Z. You, M. Tan, Diverse Roles of Mitochondria in Immune Responses: Novel Insights Into Immuno-Metabolism. *Front. Immunol.* **9**, 1605 (2018).

681. P. Schönfeld, L. Wojtczak, Short- and medium-chain fatty acids in energy metabolism: the cellular perspective. *J. Lipid Res.* **57**, 943–954 (2016).

682. A. F C Lopes, Mitochondrial metabolism and DNA methylation: a review of the interaction between two genomes. *Clin. Epigenetics*. **12**, 182 (2020).

683. J. J. Brocks, B. J. Nettersheim, P. Adam, P. Schaeffer, A. J. M. Jarrett, N. Güneli, T. Liyanage, L. M. van Maldegem, C. Hallmann, J. M. Hope, Lost world of complex life and the late rise of the eukaryotic crown. *Nature*. **618**, 767–773 (2023).

684. K. Müntjes, S. K. Devan, A. S. Reichert, M. Feldbrügge, Linking transport and translation of mRNAs with endosomes and mitochondria. *EMBO Rep.* **22**, e52445 (2021).

685. Y. Reizel, A. Spiro, O. Sabag, Y. Skversky, M. Hecht, I. Keshet, B. P. Berman, H. Cedar, Gender-specific postnatal demethylation and establishment of epigenetic memory. *Genes Dev.* **29**, 923–933 (2015).

686. A. Sanyal, B. R. Lajoie, G. Jain, J. Dekker, The long-range interaction landscape of gene promoters. *Nature*. **489**, 109–113 (2012).

687. B. Borsari, P. Villegas-Mirón, S. Pérez-Lluch, I. Turpin, H. Laayouni, A. Segarra-Casas, J. Bertranpetit, R. Guigó, S. Acosta, Enhancers with tissue-specific activity are enriched in intronic regions. *Genome Res.* **31**, 1325–1336 (2021).

688. M. Khandekar, W. Brandt, Y. Zhou, S. Dagenais, T. W. Glover, N. Suzuki, R. Shimizu, M. Yamamoto, K.-C. Lim, J. D. Engel, A Gata2 intronic enhancer confers its pan-endothelia-specific regulation. *Development*. **134**, 1703–1712 (2007).

689. C. J. Ott, N. P. Blackledge, J. L. Kerschner, S.-H. Leir, G. E. Crawford, C. U. Cotton, A. Harris,

Intronic enhancers coordinate epithelial-specific looping of the active CFTR locus. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 19934–19939 (2009).

690. S. Kawase, T. Imai, C. Miyauchi-Hara, K. Yaguchi, Y. Nishimoto, S.-I. Fukami, Y. Matsuzaki, A. Miyawaki, S. Itohara, H. Okano, Identification of a novel intronic enhancer responsible for the transcriptional regulation of musashi1 in neural stem/progenitor cells. *Mol. Brain.* **4**, 14 (2011).

691. A. K. Maunakea, R. P. Nagarajan, M. Bilenky, T. J. Ballinger, C. D'Souza, S. D. Fouse, B. E. Johnson, C. Hong, C. Nielsen, Y. Zhao, G. Turecki, A. Delaney, R. Varhol, N. Thiessen, K. Shchors, V. M. Heine, D. H. Rowitch, X. Xing, C. Fiore, M. Schillebeeckx, S. J. M. Jones, D. Haussler, M. A. Marra, M. Hirst, T. Wang, J. F. Costello, Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature.* **466**, 253–257 (2010).

692. E. Batsché, J. Yi, O. Mauger, E. Kornobis, B. Hopkins, C. Hanmer-Lloyd, C. Muchardt, CD44 alternative splicing senses intragenic DNA methylation in tumors via direct and indirect mechanisms. *Nucleic Acids Res.* **49**, 6213–6237 (2021).

693. N. Ignatiadis, B. Klaus, J. B. Zaugg, W. Huber, Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nat. Methods.* **13**, 577–580 (2016).

694. T. Wu, E. Hu, S. Xu, M. Chen, P. Guo, Z. Dai, T. Feng, L. Zhou, W. Tang, L. Zhan, X. Fu, S. Liu, X. Bo, G. Yu, clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb).* **2**, 100141 (2021).

695. S. Durinck, Y. Moreau, A. Kasprzyk, S. Davis, B. De Moor, A. Brazma, W. Huber, BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics.* **21**, 3439–3440 (2005).

696. S. Durinck, P. T. Spellman, E. Birney, W. Huber, Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).

697. Q. Wang, G. M. Garrity, J. M. Tiedje, J. R. Cole, Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007).

698. M. J. Anderson, D. C. I. Walsh, PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing? *Ecol. Monogr.* **83**, 557–574 (2013).

699. E. Reisinger, L. Genthner, J. Kerssemakers, P. Kensche, S. Borufka, A. Jugold, A. Kling, M. Prinz, I. Scholz, G. Zipprich, R. Eils, C. Lawerenz, J. Eils, OTP: An automatized system for managing and processing NGS data. *J. Biotechnol.* **261**, 53–62 (2017).

700. F. Mölder, K. P. Jablonski, B. Letcher, M. B. Hall, C. H. Tomkins-Tinch, V. Sochat, J. Forster, S. Lee, S. O. Twardziok, A. Kanitz, A. Wilm, M. Holtgrewe, S. Rahmann, S. Nahnsen, J. Köster, Sustainable data analysis with Snakemake. *F1000Res.* **10**, 33 (2021).

701. W. Huber, V. J. Carey, R. Gentleman, S. Anders, M. Carlson, B. S. Carvalho, H. C. Bravo, S. Davis, L. Gatto, T. Girke, R. Gottardo, F. Hahne, K. D. Hansen, R. A. Irizarry, M. Lawrence, M. I. Love, J. MacDonald, V. Obenchain, A. K. Oleś, H. Pagès, A. Reyes, P. Shannon, G. K. Smyth, D. Tenenbaum, L. Waldron, M. Morgan, Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods.* **12**, 115–121 (2015).

702. A. Mayakonda, M. Schönung, J. Hey, R. N. Batra, C. Feuerstein-Akgoz, K. Köhler, D. B. Lipka, R. Sotillo, C. Plass, P. Lutsik, R. Toth, Methrix: an R/Bioconductor package for systematic

aggregation and analysis of bisulfite sequencing data. *Bioinformatics*. **36**, 5524–5525 (2021).

703. K. Korthauer, S. Chakraborty, Y. Benjamini, R. A. Irizarry, Detection and accurate false discovery rate control of differentially methylated regions from whole genome bisulfite sequencing. *Biostatistics*. **20**, 367–383 (2019).

704. S. Ramchandani, S. K. Bhattacharya, N. Cervoni, M. Szyf, DNA methylation is a reversible biological signal. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 6107–6112 (1999).

705. A. J. Bannister, T. Kouzarides, Reversing histone methylation. *Nature*. **436**, 1103–1106 (2005).

706. R. R. Rodrigues, M. Gurung, Z. Li, M. García-Jaramillo, R. Greer, C. Gaulke, F. Bauchinger, H. You, J. W. Pederson, S. Vasquez-Perez, K. D. White, B. Frink, B. Philmus, D. B. Jump, G. Trinchieri, D. Berry, T. J. Sharpton, A. Dzutsev, A. Morgun, N. Shulzhenko, Transkingdom interactions between Lactobacilli and hepatic mitochondria attenuate western diet-induced diabetes. *Nat. Commun.* **12**, 101 (2021).

707. A. Franco-Obregón, J. A. Gilbert, The Microbiome-Mitochondrion Connection: Common Ancestries, Common Mechanisms, Common Goals. *mSystems*. **2** (2017), doi:10.1128/mSystems.00018-17.

708. V. Lloréns-Rico, J. Cano, T. Kamminga, R. Gil, A. Latorre, W.-H. Chen, P. Bork, J. I. Glass, L. Serrano, M. Lluch-Senar, Bacterial antisense RNAs are mainly the product of transcriptional noise. *Sci Adv*. **2**, e1501363 (2016).