



UNIVERSITAT
ROVIRA i VIRGILI

ESTUDI DE LES ALTERACIONS METABÒLIQUES EN NENS I NENES
DEGUDES A L'EXPOSICIÓ AMBIENTAL

Cèlia Torres Vilanova

TREBALL FINAL DE GRAU BIOTECNOLOGIA

Tutor acadèmic: Javier Capilla Luque, investigador i professor titular, Ciències
Mèdiques Bàsiques, javier.capilla@urv.cat

En cooperació amb: Institut d'Investigació Sanitària Pere i Virgili

Supervisora: Noelia Ramírez González, Investigadora vinculada direcció tesis, Eng.
Electrònica, Elèctrica i Automàtica, noelia.ramirez@urv.cat

Juny 2022

Jo, Cèlia Torres Vilanova, amb DNI 39470583-F, sóc coneixedora de la guia de prevenció del plagi a la URV Prevenció, detecció i tractament del plagi en la docència: guia per a estudiants (aprovada el juliol 2017) (<http://www.urv.cat/ca/vidacampus/serveis/crai/que-us-oferim/formacio-competencies-nuclears/plagi/>) i afirmo que aquest TFG no constitueixen cap de les conductes considerades com a plagi per la URV.

Tarragona, 07 de Juny 2022

A handwritten signature in black ink, consisting of a large, stylized 'C' with a horizontal line extending to the right and a vertical line extending downwards from the center of the 'C'.

(signatura)

ÍNDEX

1. DADES DEL CENTRE.....	5
2. RESUM I PARAULES CLAU.....	5
3. INTRODUCCIÓ	6
3.1. Exposició passiva al fum del tabac	6
3.1.1. Efectes adversos de l'exposició a tòxics del tabac en nens i nenes	7
3.2. Metabolòmica.....	8
3.2.1. Flux de treball metabolòmica no dirigida.....	9
4. HIPÒTESIS I OBJECTIUS	10
5. METODOLOGIA.....	10
5.2. Informació sobre la cohort.....	10
5.3. Recollida de les mostres d'orina	10
5.4. Extracció de metabòlits	11
5.5. Anàlisi metabolòmica	12
5.5.1. HERMES	12
5.6. Processament de dades en metabolòmica	14
5.6.1. XCMS	14
5.6.2. Filtrat de dades de metabolòmica	16
6. RESULTATS, DISCUSSIÓ I RELACIÓ AMB ELS OBJECTIUS PLANEJATS ..	17
6.1. Anàlisi de la qualitat de les dades adquirides	17
6.2. Espectres de MS ² HERMES	19
6.3. Optimització dels paràmetres del processat en metabolòmica (XCMS)	19
6.4. XCMS.....	19
6.5. Filtrat de dades de metabolòmica	20
6.6. Identificació de biomarcadors d'exposició al fum del tabac.....	22
7. CONCLUSIONS.....	24
8. PERSPECTIVES DE FUTUR.....	¡Error! Marcador no definido.
9. BIBLIOGRAFIA	25
10. AUTOAVALUACIÓ.....	28
11. ANNEXOS.....	28
11.01. Metadades exposició al fum del tabac	28
11.02. Metadades ingesta aliments que alteren el metaboloma	31
11.03. Metadades urianàlisi	32
11.04. Integració de pics i %RSD de la nicotina-d3	34
11.05. Integració de pics i %RSD de la cotinina-d3	36
11.06. Codi llenguatge R Total Ion Chromatogram (TIC).....	39

11.07. Codi de llenguatge R Extracted Ion Chromatogram (EIC)	40
11.08. Base de dades THS	40
11.09. Codi de llenguatge R de XCMS paràmetres optimitzats	45
11.10. Codi de llenguatge R filtrat de les dades de metabolòmica	46
11.11. Codi de llenguatge R del boxplot nicotina i hidroxicotinina	48

1. DADES DEL CENTRE

He dut a terme el treball de final de grau al grup Metabolomics Interdisciplinary Laboratory (MIL@b) que pertany al Departament d'Enginyeria Elèctrica, Elèctrica i Automàtica de la Universitat Rovira i Virgili (URV) i a l'Institut d'Investigació Sanitària Pere i Virgili (IISPV). L'objectiu del grup de recerca és avançar en el coneixement científic i tecnològic en metabolòmica, ciència que es dedica a l'estudi global dels composts químics de baix pes molecular que es troben en biofluids i teixits dels organismes vius.

La meva estada dins del grup MIL@b es duu a terme a la línia de recerca Toxicology & Environmental Metabolomics, liderada per la Doctora Noelia Ramírez, la meva tutora professional. El meu treball de final de grau es part del projecte ECHOES (*Epigenetic and metabolic alterations associated to early childhood exposure to air pollution and its role in later life chronic disease*) finançat per la Fundació La Caixa amb codi LCF/PR/SR19/52540002. L'objectiu del projecte és investigar les possibles desigualtats socials dels nens respecte a la seva exposició a la contaminació de l'aire (centrant els estudis en l'exposició passiva al fum del tabac), el paper d'aquesta exposició en la salut dels nens i el seu possible impacte en el desenvolupament de malalties cròniques en el seu futur.

2. RESUM I PARAULES CLAU

Introducció: El tabaquisme passiu causa aproximadament l'1% de la mortalitat mundial. La determinació de biomarcadors del fum del tabac en matrius biològiques com l'orina és clau per avaluar els efectes sobre la salut que pot causar aquesta exposició. Hi ha una manca d'estudis que caracteritzin les alteracions metabòliques i epigenètiques induïdes per l'exposició passiva al fum del tabac en nens.

Objectius: L'objectiu principal d'aquest treball és analitzar mostres d'orina de nens i nenes de quatre anys per determinar les possibles alteracions metabòliques entre els menors que hi conviuen amb fumadors i els que no hi conviuen. L'objectiu específic d'aquest treball és el d'aprendre el flux de treball de metabolòmica no dirigida en orines, des de l'extracció dels metabòlits a l'anàlisi dels extractes per cromatografia acoblada a espectrometria de masses i el processat de les dades de metabolòmica.

Materials i mètodes: Estudi de metabolòmica no dirigida en què s'analitzen mostres d'orina de nens de 4 anys mitjançant cromatografia líquida acoblada a espectrometria de masses. Les dades de metabolòmica es processen mitjançant XCMS. Per identificar els composts és necessari fer un MS² (MS/MS) dels ions biològicament més rellevants, aquests s'escullen amb l'eina HERMES.

Resultats: La intensitat relativa de la nicotina és més alta en nens que conviuen amb fumadors que en nens que no hi conviuen, però la presència ubíqua de nicotina confirma que l'exposició passiva al fum del tabac està present per igual independentment de la convivència amb fumadors.

Conclusió: Tot i que la metabolòmica no dirigida té inconvenients com l'obtenció de conjunts de dades molt complexos, la seva proximitat al fenotip permet identificar alteracions metabòliques entre els grups d'exposats i no exposats al fum del tabac.

Paraules clau: Exposició ambiental en població pediàtrica, orina, metabolòmica no dirigida, processat de dades en llenguatge R

3. INTRODUCCIÓ

3.1. Exposició passiva al fum del tabac

L'anomenat *secondhand smoke* (SHS, en anglès), és una barreja complexa formada pel fum del tabac exhalat pel fumador/a i el fum emès pel tabac en combustió ¹. La principal via d'exposició a SHS és la inhalació i el fum pot perdurar a l'ambient de minuts a hores ².

L'agència internacional per a la recerca del càncer (IARC) classifica 63 substàncies químiques reportades del fum del tabac com cancerígenes, 11 d'elles com a carcinògens humans, amb un paper provat en el desenvolupament de diferents tipus de càncer, malalties cardiovasculars i ictus ^{3,4}.

El fum residual del tabac, també conegut com a *thirdhand smoke* (THS), és la contaminació que persisteix després que el fum del tabac s'hagi emès a l'aire. El THS està compost pels gasos i les partícules del fum del tabac que s'adhereixen als materials, a les superfícies i a les partícules de la pols. Aquests components també poden reaccionar amb els oxidants atmosfèrics per produir subproductes potencialment nocius, que es poden tornar a emetre a la fase gasosa. Les partícules de l'aire que inicialment es van dipositar a les superfícies interiors, es poden tornar a

suspendre fins mesos després que acabi el tabaquisme actiu. Les vies d'exposició a THS són: la inhalació, absorció dèrmica de superfícies contaminades i ingestió del THS que s'adhereix a les mans o als aliments ².

Actualment, els perills associats al THS en humans no estan tan ben documentats com els del SHS. Tot i això, en models de ratolí s'ha observat que l'exposició a THS durant un mes, provoca un augment de les citocines proinflamatòries i de l'aspartat aminotransferasa, un biomarcador de dany hepàtic. També es va afirmar que l'exposició durant 4-6 mesos alterava molts factors relacionats amb l'estrès oxidatiu, causant danys a nivell molecular. Es va veure que els ratolins presentaven hiperglucèmia i hiperinsulinèmia, fet que suggereix que l'exposició durant un llarg període a THS pot provocar resistència a la insulina ⁵.

Malgrat les evidències emergents sobre la toxicitat i carcinogenicitat del THS, aquesta forma d'exposició als contaminants del fum del tabac encara no és reconeguda per la majoria de la població i s'ha omès en les polítiques de salut pública i mediambiental ¹.

3.1.1. Efectes adversos de l'exposició a tòxics del tabac en nens i nenes

La legislació sobre l'aire interior net i lliure de fum ha reduït l'exposició del fum ambiental del tabac a llocs públics ⁶, i per això l'entorn domèstic s'ha convertit en la principal font d'exposició al tabaquisme passiu. Un estudi recent a Espanya (n=2411), demostra que més del 29% dels nens d'entre 6 i 11 anys està exposat a SHS a casa seva o al cotxe i més del 42% continua exposat a llocs públics ⁷. Aquestes dades subestimen l'exposició dels nens al THS, una font d'exposició menys estudiada que afecta sobretot als nens que conviuen amb fumadors.

La pols i les superfícies domèstiques són fonts importants d'exposició a THS en nens petits, a causa de comportaments associats a l'edat i les interaccions amb el seu entorn (ex. contacte físic proper amb els pares, posar-se la mà a la boca, etc.) ^{2,4}. S'estima que els nens de cinc anys, que passen la major part del temps en llocs interiors, ingereixen una mitjana de 100 mg de pols interior al dia ⁸.

Els nens són especialment vulnerables durant el desenvolupament fetal i els seus primers anys de vida a l'exposició a SHS i THS, perquè els pulmons, òrgans i cervell encara estan madurant. D'altra banda, els nens tenen un volum pulmonar més petit, i per això respiren més ràpidament, inhalant així més contaminants ^{4,9}.

El SHS està relacionat amb conseqüències greus per a la salut infantil: infeccions d'òida, síndrome de mort sobtada del nadó, trastorns neuroconductuals i símptomes i infeccions respiratòries (asma, tos, dificultat per respirar, bronquitis, etc.), que van provocar 543000 morts de nens l'any 2016 ^{4,10}.

Recentment, s'ha dut a terme un estudi amb nens de 5 anys (n=25), que es presentaven a urgències amb una simptomatologia potencialment relacionada amb SHS. Es van avaluar els nivells de nicotina a les mans i de cotinina salival. Tots els nens van presentar nivells detectables de nicotina a les mans (18,3-690,9 ng) i tots menys un van presentar nivells detectables de cotinina salival (1,2-28,8 ng/mL) ¹¹.

3.2. Metabolòmica

La metabolòmica es centra en el perfilat i la quantificació a gran escala de metabòlits presents a les cèl·lules, teixits, biofluids o organismes. El conjunt de metabòlits i les seves interaccions dins d'un sistema biològic constitueixen el metaboloma ¹². El metaboloma és molt dinàmic: els metabòlits s'absorbeixen, sintetitzen, degraden i interaccionen contínuament dins i entre sistemes biològics i amb el medi ambient. Aquestes reaccions formen vies metabòliques interconnectades que constitueixen una xarxa metabòlica.

La metabolòmica és la ciència òmica més pròxima al fenotip, ja que els metabòlits reflecteixen directament l'activitat bioquímica subjacent i l'estat de les cèl·lules/teixits, i per això és ideal per detectar perturbacions en el sistema biològic ¹².

Donat el gran nombre de metabòlits amb diversitat química i estructural, la complexitat de les matrius biològiques i les baixes concentracions dels metabòlits d'interès és crucial el desenvolupament de mètodes analítics d'alt rendiment, robusts i sensibles. La tecnologia més utilitzada en metabolòmica és l'espectrometria de masses d'ionització per *electrospray* acoblada a cromatografia líquida (LC-ESI-MS) ¹³.

Hi ha dos enfocaments basats en espectrometria de masses: metabolòmica dirigida i no dirigida. S'usa metabolòmica dirigida quan tenim una hipòtesi inicial i busquem un grup petit de metabòlits. En canvi, la metabolòmica no dirigida s'usa per a perfilar el màxim nombre de metabòlits possible i generar una hipòtesi ¹³.

3.2.1. Flux de treball metabolòmica no dirigida

En aquest estudi de metabolòmica no dirigida, donat el gran nombre de metabòlits amb diversitat química i estructural que constitueixen el metaboloma, s'han de tenir en compte un gran nombre de factors per poder obtenir resultats fiables. Aquests factors inclouen la preparació i l'anàlisi de les mostres, el processament de les dades i la identificació dels espectres MS/MS.

1. **Preparació de les mostres:** S'han d'utilitzar mètodes d'extracció que permetin detectar la major quantitat de metabòlits simultàniament. És important que els extractes siguin compatibles amb la tècnica analítica usada.
2. **Anàlisi de les mostres:** En aquest estudi, l'anàlisi de les mostres es duu a terme mitjançant cromatografia líquida acoblada a un espectròmetre de masses (LC-ESI-MS). La cromatografia líquida (LC) es basa en la interacció dels metabòlits amb la fase estacionària de la columna cromatogràfica. Depenent de les propietats químiques dels metabòlits, aquests es separaran i eluiran en un temps de retenció específic. Abans que els metabòlits passin a l'espectròmetre de masses s'han d'ionitzar. La tècnica d'ionització ESI és una tècnica d'ionització suau que tendeix a conservar els ions moleculars intactes mitjançant la formació d'adductes. L'espectrometria de masses (MS) mesura la relació massa/càrrega i la intensitat dels composts. Per identificar els composts són necessaris mètodes d'adquisició MS/MS. Per identificar els ions per fer MS/MS fem servir l'eina HERMES, que és una tècnica basada en *suspect screening* que consisteix en la detecció de llistes personalitzades de composts químics d'interès (que sospitem que estiguin a les mostres) utilitzant la seva fórmula molecular i la massa exacta resultant ¹⁴.
3. **Processament de les dades:** Usem XCMS, que ens proporciona mètodes per poder detectar pics cromatogràfics i alinear-los entre les mostres. Aquest alineament permet que el software pugui agrupar pics per obtenir *features*. Una *feature* és un conjunt de dades de relacions massa/càrrega en un temps de retenció determinat. Generalment, un estudi LC-ESI-MS proporciona milers de *features*, que s'han de filtrar per assolir resultats fiables.
4. **Identificació espectres MS/MS:** Consisteix a comparar espectres obtinguts de l'anàlisi MS/MS de les mostres amb espectres de bases de dades com la NIST o METLIN.

4. HIPÒTESIS I OBJECTIUS

La principal hipòtesis d'aquest estudi és que els composts tòxics del fum del tabac són contaminants ambientals ubics als quals estan exposats tant els infants que conviuen amb fumadors com els que no hi conviuen. D'altra banda, aquesta exposició pot produir alteracions en el metaboloma dels infants, especialment d'aquells més exposats al fum del tabac.

L'objectiu principal d'aquest treball és analitzar mostres d'orina de nens i nenes de quatre anys per determinar les possibles alteracions metabòliques entre els menors que hi conviuen amb fumadors i els que no hi conviuen. L'objectiu específic d'aquest treball és el d'aprendre el flux de treball de metabòlica no dirigida en orines, des de l'extracció dels metabòlits a l'anàlisi dels extractes per cromatografia acoblada a espectrometria de masses i el processat de dades de metabòlica.

5. METODOLOGIA

5.2. Informació sobre la cohort

La cohort infantil per aquest estudi és una subcohort dels participants del projecte ECLIPSES (Ensayo Clínico para Suplementar con Hierro a Embarazadas) finançat per Fondos de Investigación Sanitaria Instituto de Salud Carlos III (PI12/02777 i PI17/01754). Tots els pares/tutors legals van signar un formulari de consentiment informat aprovat per la *Clinical Research Ethical* (Codi de referència: 155/2017).

Els participants d'aquest projecte són 146 nens i nenes de 4 anys, residents en diferents poblacions de la província de Tarragona. Dels 146 parcipants, 51 conviuen amb fumadors i 95 amb no fumadors. Els participants van proporcionar una mostra de la primera orina del dia. Els adults responsables van respondre un qüestionari sobre l'exposició a fum del tabac (convivència amb persones fumadores, fumar en presència dels menors i nombre de cigarretes que fumen al dia) [Annex 11.01]. Es va informar els adults responsables què els menors no ingerissin xocolata, peix ni begudes amb gas el dia anterior, ja que aquestes alteren el metaboloma.

5.3. Recollida de les mostres d'orina

Durant la recollida de les mostres, els adults responsables van informar si els menors havien ingerit xocolata, peix o begudes amb gas el dia anterior [Annex 11.02]. La recollida de les mostres es va dur a terme en una nevera portàtil. Les mostres es van centrifugar i es van guardar en alíquotes d'1mL a -80°C el mateix dia de la recollida.

Abans d'emmagatzemar-les, es va fer un urinàlisi per detectar el nivell de leucòcits, la presència d'hemoglobina en sang i la densitat relativa de les mostres, entre altres paràmetres [Annex 11.03.].

5.4. Extracció de metabòlits

El protocol d'extracció de metabòlits de les mostres d'orina [figura 1] és una adaptació del protocol descrit per Southam et al. ¹⁶. L'extracció es va realitzar a partir de 50µL d'orina (n=146) als quals es van afegir 150 µL de solució extractant [Figura 2]. Els extractes es van centrifugar a 15000 rpm durant 10 minuts a 4 °C. Es van fer alíquotes del sobrenedant per al seu posterior anàlisi per cromatografia líquida d'alta pressió acoblada a espectrometria de masses d'alta resolució (UHPLC-HRMS).

Es van preparar blancs d'extracció, on es substitueixen els 50µL d'orina per 50µL d'aigua MilliQ. Es van preparar controls de qualitat (QC) de totes les mostres fent una mescla de 5µL de cada mostra d'orina. Els QCs s'utilitzen per controlar la qualitat de l'adquisició de dades i QC_exposed de nens exposats al fum del tabac.

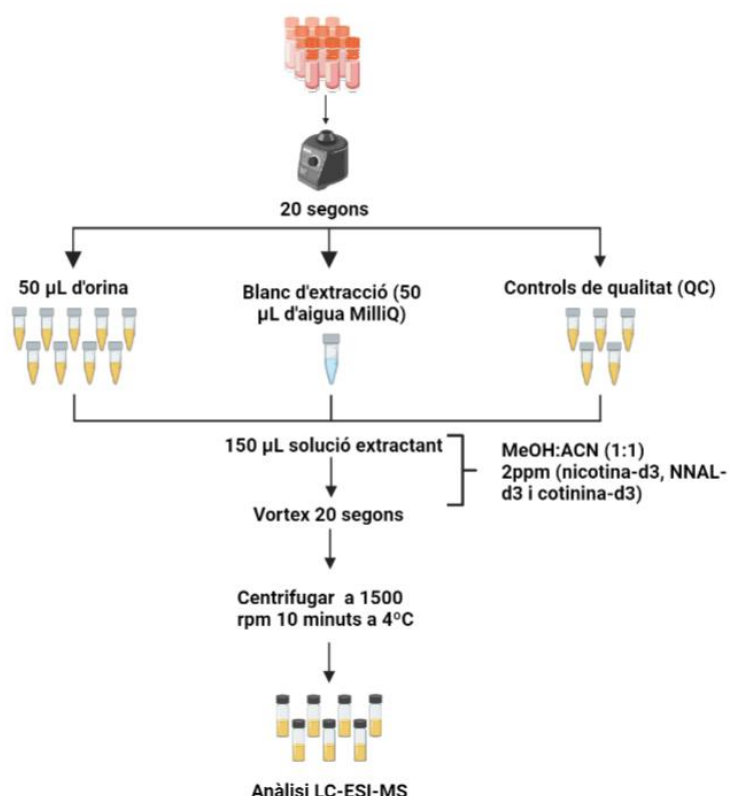


Figura 1. Les mostres d'orina es mesclen al vòrtex durant 20 segons. S'afegeixen 150µL d'extractant a 50µL d'orina (En el cas dels blancs d'extracció, els 50µL d'orina es substitueixen per 50µL d'aigua MilliQ). Es mesclen al vòrtex durant 20 segons, es centrifuguen i s'analitzen a l'equip LC-ESI-MS; Font: Figura pròpia realitzada amb Biorender (<https://biorender.com/>)

5.5. Anàlisi metabolòmica

S'injecten 2µL de cada mostra al cromatògraf líquid d'alta resolució (Vanquish Flex UHPLC) acoblat a l'espectròmetre de masses (Orbitrap ID-X™ Tribrid™) en mode d'ionització *electrospray* positiu (ESI+).

Els metabòlits es separen amb la columna HILIC (UPLC BEH HILIC 1.7µm (150 x 2.1mm). La fase mòbil està composta per A= acetonitril i B= Acetat d'amoni 50 mM en aigua ultrapura, amb un flux de 0,4m/min. Les condicions ESI+ van ser 3,5kV, temperatura del gas 300°C.

Aquest experiment LC-ESI-MS, genera milions de punts de dades tridimensionals (m/z, temps de retenció, intensitat) que mitjançant softwares com XCMS s'acabaran convertint en milions de *features*. Per poder identificar els metabòlits són necessaris mètodes d'adquisició MS/MS. Per determinar els ions rellevants biològicament per fer MS/MS utilitzem HERMES ¹⁷, desenvolupat al grup de recerca MIL@b.

5.5.1. HERMES

Els mètodes convencionals per fer MS/MS com el DDA (Data-dependent acquisition), prioritza fragmentar els ions més abundants, però no necessàriament els que són més rellevants biològicament ¹⁷. Per aquest motiu, en aquest estudi fem servir el mètode experimental HERMES, que prioritza els ions rellevants biològicament.

HERMES és un mètode que interroga directament als punts tridimensionals obtinguts en els experiments LC-ESI-MS mitjançant una llista complexa de fórmules moleculars úniques que l'usuari pot personalitzar. Cada fórmula molecular genera múltiples fórmules iòniques afegint o restant àtoms dels adductes més comuns.

HERMES és una eina computacional que està disponible com a paquet de R (RHermes). El flux de treball d'HERMES [figura 2] consisteix en:

1. L'usuari escull una base de dades d'interès que contingui fórmules moleculars, també pot personalitzar la llista d'adductes d'interès. HERMES amb la llista d'adductes i les fórmules moleculars, generarà una llista de fórmules iòniques.
2. Els punts de dades tridimensionals de LC/MS s'interroguen amb la llista de fórmules iòniques per fer anotacions de MS1.

3. Els punts amb les mateixes anotacions m/z s'agrupen per densitat en un interval de temps de retenció específic formant els *scans of interest* (SOI). Com a pas addicional es pot extreure el blanc, de manera que els SOIs amb forma i intensitat similar al blanc s'eliminen
4. Els SOIs corresponents a diferents adductes de la mateixa fórmula molecular s'agrupen mitjançant el seu perfil d'elució (Si els adductes provenen de la mateixa fórmula molecular tindran un perfil d'elució molt similar)
5. Aquestes SOIs es poden filtrar per acabar obtenint la llista d'inclusió (IL). La IL estarà formada per fragments específics de la mostra i no redundants, i serà aquesta llista d'inclusió de la qual farem els MS/MS.

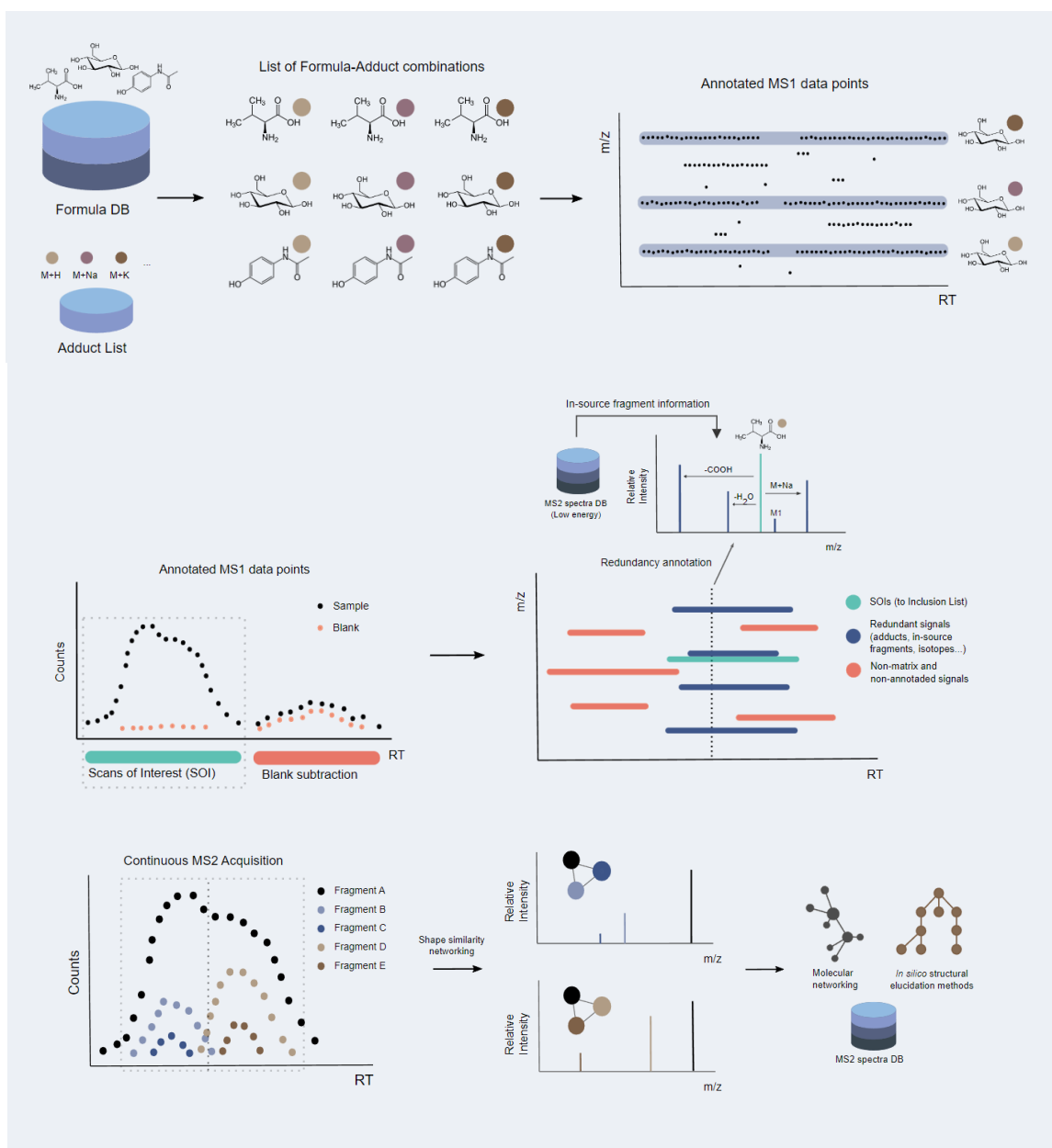


Figura 2. Resum gràfic del flux de treball de HERMES; Font: Giné et al, 2021

En aquest estudi utilitzem tres bases de dades: la THS (una llista pròpia de 177 composts específics del tabac), una part de la Human Metabolome Database (composada per 4161 metabòlits detectats en orina humana), i la CEC (composada per 283501 composts tòxics relatius a l'exposoma humà).



Figura 3. En aquest estudi, quan fem l'anàlisi de les mostres obtenim MS¹. Per poder identificar els metabòlits necessitem espectres MS² (MS/MS), i per això fem servir l'eina HERMES. HERMES interroga a les dades tridimensionals de l'anàlisi amb les tres bases de dades que hem personalitzat (THS, HMDB i CEC). HERMES acabarà generant una llista d'inclusió amb els fragments més rellevants biològicament, que serà els que analitzarem per obtenir MS². Font pròpia: Biorender (<https://biorender.com/>)

HERMES caracteritza les mostres, però no ens dona intensitats relatives dels composts, per això fem servir XCMS, que permet fer comparacions sobre les diferències en la proporció de metabòlits entre mostres de nens exposats i no exposats

5.6. Processament de dades en metabolòmica

5.6.1. XCMS

En aquest estudi, el processament de les dades MS¹ es duu a terme amb XCMS. XCMS proporciona mètodes per detectar pics, alinear-los entre mostres i agrupar els pics que s'han alineat entre les mostres per formar les *features*, les quals quantifica de manera relativa. El més important en el processament de XCMS són els paràmetres a utilitzar, ja que uns paràmetres que no s'ajustin a les dades adquirides poden donar lloc a resultats esbiaixats. És important comprovar que els paràmetres establerts propis de l'equip a analitzar funcionen bé amb els pics models, en el nostre cas, els estàndards interns.

XCMS online proporciona els paràmetres per defecte del nostre equip [figura 4]. De manera experimental, inspeccionem les dades crues amb el software comercial de l'equip (Xcalibur). Calculem l'error de massa (ppm) dels estàndards interns respecte a la massa monoisotòpica protonada teòrica i observem que l'error de massa és superior, i per això establim un error de 5ppm. Usem la resta de paràmetres que ens proporciona xcms *online* del nostre equip per defecte.

The screenshot shows the 'Retention Time Correction' tab in the XCMS online software. The 'Method' is set to 'centWave'. The parameters and their values are as follows:

Option	Value	Note
ppm	2.5	maximal tolerated m/z deviation in consecutive scans, in ppm (parts per million)
minimum peak width	5	minimum chromatographic peak width in seconds note: must be less than max peak width. See also here .
maximum peak width	20	maximum chromatographic peak width in seconds note: must be greater than min peak width. See also here .
View Advanced Options		
mzdiff	0.01	minimum difference in m/z for peaks with overlapping retention times, can be negative to allow overlap
Signal/Noise threshold	10	Signal/Noise threshold
Integration method	1	Integration method. If =1 peak limits are found through descent on the mexican hat filtered data, if =2 the descent is done on the real data. Method 2 is very accurate but prone to noise, while method 1 is more robust to noise but less exact.
prefilter peaks	3	Prefilter step for the first phase. Mass traces are only retained if they contain at least [prefilter peaks] peaks with intensity >= [prefilter intensity]
prefilter intensity	5000	Prefilter step for the first phase. Mass traces are only retained if they contain at least [prefilter peaks] peaks with intensity >= [prefilter intensity]
Noise Filter	1000	optional argument which is useful for data that was centroided without any intensity threshold, centroids with intensity < noise are omitted from ROI detection

Figura 5. Paràmetres del nostre equip LC-ESI-MS per defecte; Font: <https://xcmsonline.scripps.edu/>

El flux de treball [figura 6] de XCMS consisteix a:

1. Detecció dels pics cromatogràfics: Utilitzem el mètode CentWave per detectar els pics, aquest mètode és molt sensible, de manera que detectem una gran quantitat de pics (inclosos pics no rellevants). Els paràmetres més importants en la detecció de pics són: **peakwidth** (amplada de pic) i els **ppm** (error de massa).
2. Alineament dels pics entre les mostres: En l'anàlisi LC-ESI-MS es produeixen desviacions en els temps de retenció (canvi en el flux, PH...), i per això és important que aquest software permeti ajustar les petites desalineacions que es puguin dur a terme en l'anàlisi. L'alineament el que fa es fer ample/llarg el cromatograma per corregir aquestes petites desalineacions.

3. Agrupació dels pics cromatogràfics: Els pics s'agruparan segons la relació massa/càrrega i el temps de retenció. Els paràmetres més importants en l'agrupació són: **bw** (defineix la suavitat de la funció, està relacionat amb l'amplada en el temps de retenció per encabir els pics) i el **binsize** (Defineix el rang de relació massa/càrrega que es necessita per fer encabir els pics). Els pics que s'alineen entre mostres i s'agrupen formen les *features*.
4. Buscar valors perduts: La detecció de pics pot fallar, de manera que retorna un valor NA, quan l'ió no està present en la mostra, també retornarà un valor NA. La funció *fillChromPeaks* permet omplir el senyal dels pics que falten de l'àrea de la *feature* dels fitxers mzML originals.



Figura 6. Flux de treball XCMS; Font pròpia: Biorender (<https://biorender.com/>)

6.5.2. Filtrat de dades de metabolòmica

La taula de *features* que obtenim amb XCMS conté moltes *features* redundants, degut a la presència de diferents adductes o isòtops del mateix metabòlit. Un metabòlit sovint està representat per més d'una *feature*. Aquest fenomen requereix un procés de filtració de les *features* per poder obtenir resultats confiables. Els criteris de filtració consisteixen a:

1. Normalitzo les meves dades per la densitat relativa: He de normalitzar a l'inici, per tal de no tenir resultats esbiaixats a causa de la concentració en les mostres d'orina. La normalització es duu a terme dividint la matriu d'intensitats de les mostres per la seva densitat relativa.
2. Establir un llindar d'intensitat mínima: Aquest pas és molt important per tal d'eliminar el soroll de l'equip. D'altra banda, una *feature* amb molt baixa intensitat sovint no es fragmenta correctament en MS/MS i no es pot identificar. En aquest estudi utilitzem un llindar d'intensitat de 40000. En aquest pas es calcula la mediana de les intensitats, en el cas que en un dels grups tinguem aquesta *feature* amb una intensitat mediana de més de 40000 no s'eliminarà.

3. Variabilitat analítica: La variabilitat analítica es mesura entre els controls de qualitat (QC) que es van injectant de manera periòdica durant l'anàlisi. S'eliminen aquelles *features* que tenen més variabilitat analítica que biològica.
4. Filtrem per presència dels composts una base de dades: Sovint tenim *features* que provenen de la columna, aquest pas ens permet saber quines de les nostres *features* s'han detectat a la base de dades i eliminar les que no s'hagin detectat.

6. RESULTATS, DISCUSSIÓ I RELACIÓ AMB ELS OBJECTIUS PLANEJATS

6.1. Anàlisi de la qualitat de les dades adquirides

Per comprovar la qualitat de les dades calculo la desviació típica relativa (RSD) dels estàndards interns. El RSD és el valor absolut del coeficient de variació. Per tal de mesurar el %RSD utilitzo el software comercial de l'equip de LC-ESI-MS que hem utilitzat per a l'anàlisi (Xcalibur).

Integrem els pics dels estàndard interns per calcular la desviació estàndard relativa (%RSD). En cromatografia líquida, un mètode reproduïble i repetible ha de tenir un coeficient de variació inferior al 20% en els estàndards interns. En el cas de la nicotina-d3, el percentatge de variació es del 19,2% i el de la cotinina-d3 del 17,6%. Aquests %RSD em garanteix la bona qualitat de les dades adquirides. Els annexos 11.04 i 11.05 contenen més informació sobre la integració dels pics i els %RSD en nicotina-d3 i cotinina-d3 respectivament.

D'altra banda, per comprovar la qualitat de les dades faig una inspecció visual del *Total Ion Chromatogram* (TIC) dels QC [figura 7]. Aquests cromatogrames segueixen un perfil metabòlic molt similar indicant una baixa deriva analítica i corroborant la qualitat de les dades adquirides durant l'anàlisi de les mostres. El codi de llenguatge R utilitzat per fer el TIC es troba a l'Annex 11.06.

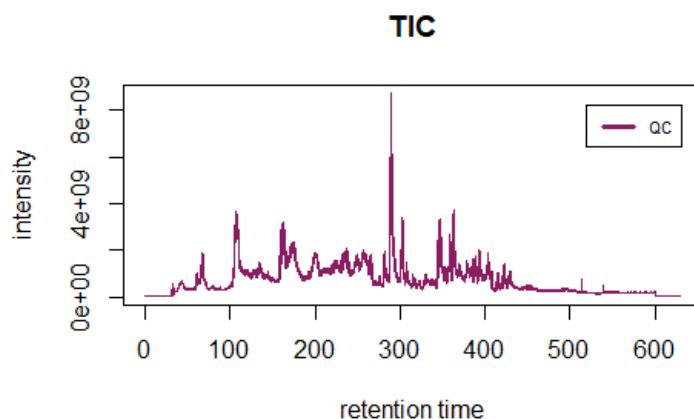


Figura 7. Total Ion Chromatogram (TIC) dels 19 QC; Font: Llenguatge de programació R

Per observar el perfil metabòlic dels estàndards interns durant l'anàlisi de les mostres, faig un *Extract Ion Chromatogram (EIC)* de la massa monoisotòpica i el temps de retenció dels estàndards als QC (El codi de llenguatge R utilitzat es troba a l'Annex 11.07). En la figura 8 observo que els perfils metabòlics dels estàndards interns en els 19 QC. El temps de retenció és més estable a l'inici de la cromatografia, i per això la nicotina-d3 és l'estàndard intern amb un temps de retenció més elevat, té més desviacions en el temps de retenció.

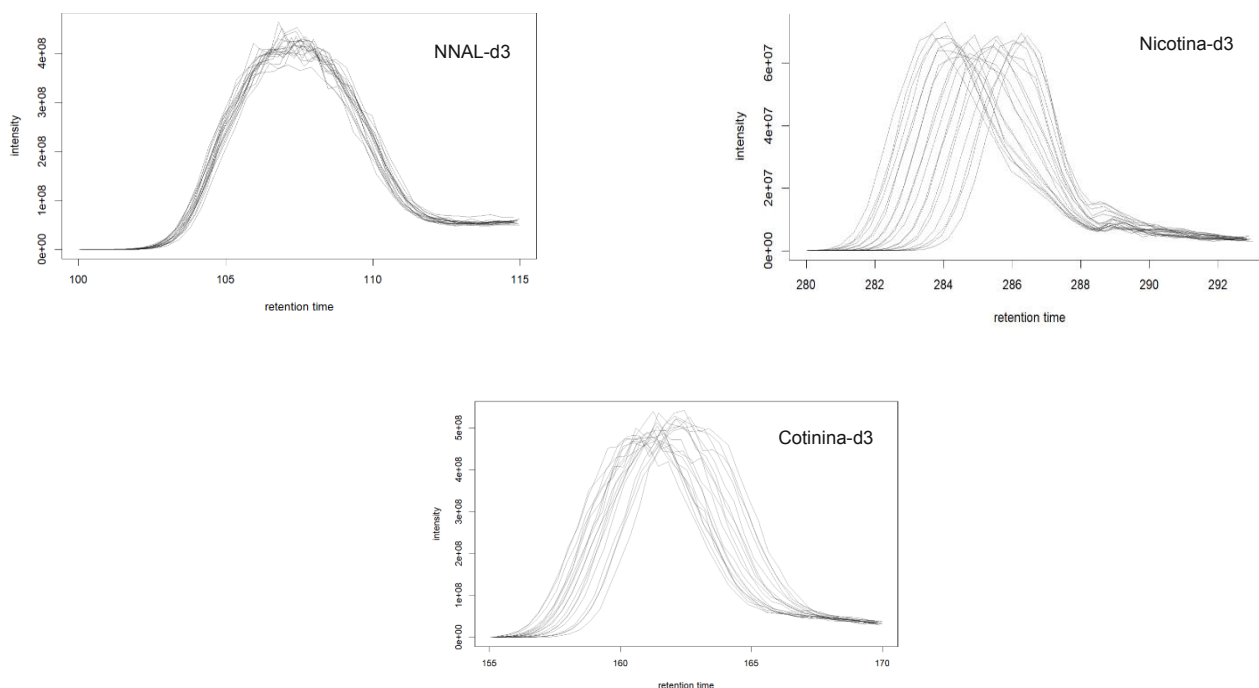


Figura 8. *Extract ion* dels estàndards interns en els QC; Font: Llenguatge de programació R

6.2. Espectres de MS² HERMES

HERMES ens dona llistes d'inclusió per fer l'anàlisi MS² dels ions biològicament rellevants. En aquest estudi, analitzem el *pool* de QC i el QC_exposats per tal de fer MS² dels fragments. Amb la base de dades anomenada THS, dels 177 composts [Annex 11.08], fem deu injeccions per tal d'obtenir MS². De la base de dades de la *Human Metabolome Database* (HMDB), dels 4161 composts, fem vint-i-cinc injeccions a l'equip per tal d'obtenir MS². De la CEC, dels 283501 composts, fem trenta injeccions a l'equip per tal d'obtenir MS².

6.3. Optimització dels paràmetres del processat en metabolòmica (XCMS)

Utilitzant els paràmetres que xcms online ens dona del nostre equip (modificant l'error ppm=2.5 per ppm=5), obtenim una matriu de *features* que agrupa pics amb desviacions de massa i desviacions del temps massa grans. Aquest fet implica que, com que els paràmetres són massa generosos, s'estan agrupant pics dins d'una mateixa *feature* que no corresponen als mateixos composts, per això els hem d'optimitzar.

El rang de masses en el qual s'agrupen les *features* ve donat pel paràmetre *binsize*. Nosaltres utilitzem un *binsize*=0.015, què ens dona una finestra de m/z massa gran, de manera que el reduïm a *binsize*=0.0005. D'altra banda, el paràmetre que determina la desviació en el temps és el *peakwidth*. Nosaltres utilitzem un *peakwidth*= c(5,20) que ens dona una finestra de temps massa gran, per això el reduïm a *peakwidth*= c(5,10). L'optimització dels paràmetres permet que la matriu de *features* sigui molt més acurada i agrupi els pics que corresponen als mateixos composts.

6.4. XCMS

El codi de llenguatge R amb els paràmetres optimitzats de XCMS es troba a l'Annex 11.09. XCMS amb els paràmetres optimitzats, detecta els pics cromatogràfics i els alinea entre les mostres per formar les *features*. Finalment, obtenim una matriu d'intensitats [Taula 1] amb 16206 *features* i 165 columnes (146 mostres + 19 QC).

	01_02_069.mzML	01_03_057.mzML	01_03_136.mzML	01_03_169.mzML	01_03_174.mzML	01_03_204.mzML	01_03_208.mzML	01_04_146.mzML	01_05_072.mzML	01_05_073.mzML
FT00001	34154.613	42557.297	33826.363	54277.79	68474.562	54594.84	NA	27842.129	NA	20238.63
FT00002	78062.383	NA	NA	NA	59752.547	77227.19	78657.461	73957.969	64638.80	75888.15
FT00003	78357.656	78403.375	68141.891	78198.04	91074.156	76103.09	80644.281	78435.438	71190.88	76010.64
FT00004	206009.703	234944.047	251240.094	191094.11	170991.953	169276.62	309189.688	243219.422	183899.88	174655.50
FT00005	1690508.625	1523596.000	1794391.500	1497778.62	894920.688	1447274.12	1367180.500	1076741.125	1372176.50	697187.19
FT00006	291703.188	328840.656	300406.438	327403.41	322289.688	315875.47	324505.344	314312.531	301881.88	305284.28
FT00007	348083.188	381837.094	342056.438	376672.69	372306.344	362695.91	364617.938	356895.375	368009.28	363103.81
FT00008	281153.562	316297.312	279133.281	308041.03	300664.062	305012.81	316914.000	304582.500	315161.44	273672.25
FT00009	296468.188	328801.438	270903.438	322877.25	304001.125	316045.69	299479.469	271460.406	319072.31	265018.34
FT00010	230585.062	277851.312	223987.562	262910.62	272148.156	256111.78	255295.812	230126.672	274344.62	224473.92
FT00011	253306.688	269792.406	227128.766	248494.50	261155.391	225779.66	253793.922	225782.938	253049.89	213912.25
FT00012	239753.844	303821.375	241299.516	260730.38	267097.812	257671.83	264941.375	248944.500	264107.66	236837.20
FT00013	11103.359	NA	NA	15283.14	13438.852	21327.26	69949.742	15972.346	18751.43	42734.65
FT00014	112347.320	202910.938	143672.391	196466.00	192562.938	110246.88	159108.359	275804.531	155225.23	404600.25

Taula 1. Matriu d'intensitats XCMS; Font: Llenguatge de programació R

6.5. Filtrat de dades de metabolòmica

En primer lloc, després de normalitzar la matriu d'intensitats per la densitat específica, fem un PCA exploratori que ens permeti fer una visualització ràpida de les nostres mostres. Com es pot observar en la figura 9, no hi ha diferències visuals entre els dos grups (exposats i no exposats al fum del tabac). D'altra banda, podem observar que els controls de qualitat (QC) estan força centrats, indicatiu d'una bona qualitat analítica.

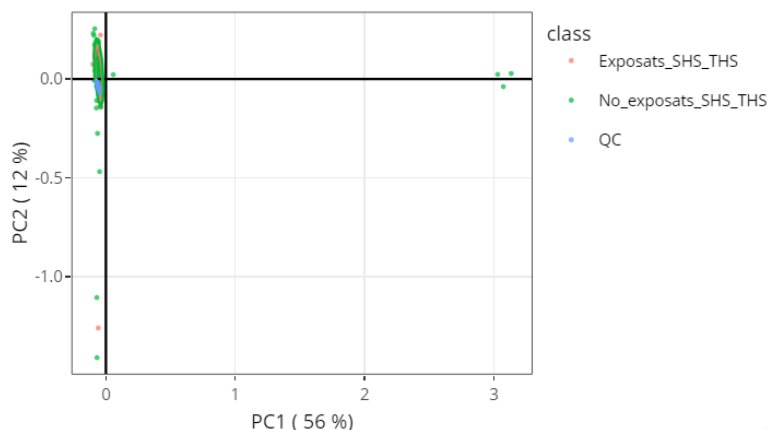


Figura 9: PCA exploratori dels nens exposats i no exposats al fum del tabac; Font: Llenguatge de programació R

Podem observar que hi ha *outliers*, ens hem d'assegurar que aquests no siguin per un problema analític, de manera que revisem les integracions dels estàndards interns i observem que en tres de les mostres (03_17_055, 03_17_003 i 04_18_003) no tenen estàndard intern, de manera que no se'ls hi va afegir solució extractant, per tant, les eliminem de l'estudi.

Procedim a filtrar les *features* per intensitat, calculem la intensitat mediana dels grups i establim un llindar mínim d'intensitat de 40000. De les 16206 *features* inicials, 11662 passen en almenys un dels grups el llindar d'intensitat [figura 10].

```

R 4.2.0 - ~/Camilla/R_files/Metadada/
> median.intensities <- as.data.frame(t(apply(D.norm,
+                                     1, function(x) tapply(x, classe.outliers, median, na.rm = TRUE))))
> ##Elimino el QC per calcular la mitjana de les intensitats
> median.intensities$QC <- NULL
> ### Estableixo un valor llindar mínim d'intensitat
> thresholdvalue <- 40000
> idx_i <- names(which(apply(median.intensities,
+                             1, function(x) any(x>thresholdvalue)==TRUE)==TRUE))
> cat(paste(length(idx_i), "out of", dim(median.intensities)[1],
+           "above the threshold intensity value"))
11662 out of 16206 above the threshold intensity value

```

Figura 10. Filtrat per intensitat; Font: Llenguatge de programació R

Procedim a eliminar les *features* que tinguin més variabilitat analítica que biològica, de manera que ens quedem amb 6667 *features* que tenen més variabilitat biològica que analítica [figura 11].

```

R 4.2.0 - ~/Camilla/R_files/Metadada/
> ## Elimino aquelles features que tinguin una major variació analítica (QC) que biològica
> idx_qc <- names(RSD_samples)[which(RSD_QC < RSD_samples)]
> cat(paste(length(idx_qc), "out of", dim(median.intensities)[1],
+           "hold higher biological than analytical variation"))
6667 out of 16206 hold higher biological than analytical variation
> |

```

Figura 11 Filtrat per variabilitat biològica > variabilitat analítica; Font: Llenguatge de programació R

Una vegada tinc aquests dos filtratges, els aplico sobre la matriu normalitzada, de manera que només em quedaré amb aquelles *features* que tinguin passin el llindar d'intensitat i tinguin una variabilitat biològica superior a la variabilitat analítica. En aquest cas, la meua matriu filtrada tindrà 4530 *features* [Taula 2].

	01_02_069	01_03_057	01_03_136	01_03_169	01_03_174	01_03_204	01_03_208	01_04_146	01_05_072	01_05_073	01_06_064	02_09_057	02_09_117	02_09_329	02_09_382	02_09_456	02_12_184
FT00028	145913.81	194465.018	124602.195	119538.460	134045.02	316186.98	15586.542	39094.09	52435.55	169848.650	1.528482e+04	4.6966.969	9533.812	13329.20	201011.04	99965.023	1.483713e+04
FT00036	394959.12	1799504.126	2942803.171	765443.537	569533.29	866941.38	126761.989	291876.30	1056614.39	414637.561	1.706138e+05	428533.303	257455.335	274800.55	652169.57	1521195.122	2.210514e+05
FT00066	NA	NA	NA	NA	NA	18071.69	NA	NA	NA	NA	NA	NA	208515.061	NA	NA	NA	NA
FT00072	214113.54	166426.881	157860.762	235678.826	219955.91	193561.97	213092.976	330014.02	68856.86	199065.807	1.124044e+05	155370.941	51468.483	176109.13	164698.49	104084.215	6.862787e+05
FT00073	354694.76	316535.255	352286.341	318053.476	305405.43	417983.31	92797.765	2678665.29	253138.98	269618.689	2.730322e+05	263762.469	158418.140	5153043.14	460093.93	240788.826	2.062402e+05
FT00102	36307992.20	5939464.563	960739.573	35456671.220	2509420.00	3306722.82	1041221.535	1978400.12	305173.84	817976.638	8.491121e+05	1663062.990	244880.732	9627446.08	2668311.46	308508.537	3.630052e+05
FT00104	1918504.02	599754.126	240431.265	1123359.146	403945.40	262103.94	96892.474	188630.43	232006.01	370971.056	9.289410e+04	365763.909	110491.082	751043.44	592144.70	79820.640	1.586157e+05
FT00123	97449.76	97941.763	339592.043	91180.534	117092.64	85654.15	31305.227	58970.10	127621.44	74411.120	1.006675e+05	49955.285	23927.323	148286.04	52937.84	59911.246	3.574035e+04
FT00135	2243574.39	313340.261	42602.744	2086311.220	120563.92	152224.44	43967.269	91119.58	NA	42056.015	4.785710e+04	66004.006	26419.266	513698.62	129448.29	15195.168	1.849851e+05
FT00143	155193.35	92348.574	66477.485	1601202.439	227355.84	151795.66	11575.286	782035.13	475795.37	2118487.864	1.006243e+05	39150.900	159076.845	56389.49	107117.08	517433.598	7.095412e+04
FT00155	189735.24	165438.987	NA	211312.896	17981.90	61907.90	29007.519	157741.63	114432.93	82335.831	1.542656e+05	NA	58620.267	NA	NA	29562.252	2.281651e+04
FT00177	NA	NA	NA	26717.788	NA	NA	7708.079	1682536.53	NA	20372.222	5.000129e+04	11029.840	NA	3073656.86	54638.38	19333.264	1.017193e+04
FT00178	79171.13	60635.278	79875.282	75161.502	84666.59	89976.81	90247.803	113870.63	32622.01	87897.277	3.592544e+04	65615.020	27200.692	91221.62	89605.92	40731.707	3.357247e+04
FT00180	106575.80	101882.494	44057.172	35064.318	73512.25	NA	11101.273	35575.89	56658.04	114792.605	1.273022e+04	33209.195	35578.438	43801.06	236031.84	22879.611	7.482648e+04
FT00181	102120.00	88379.816	44385.926	39204.825	26613.77	112878.28	10507.803	30328.74	38920.39	7236.612	9.024201e+04	8763.412	50539.790	46134.06	12046.39	22699.594	2.448613e+04
FT00194	NA	NA	NA	53802.153	349748.17	62806.46	NA	NA	65321.24	NA	NA	NA	NA	NA	NA	NA	1.950939e+04
FT00195	40154.61	59676.790	NA	32942.763	57017.13	NA	NA	141027.94	NA	53326.919	NA	127985.034	NA	45786.63	63029.13	40143.868	1.400978e+04
FT00204	225826.30	176244.342	163625.762	241599.954	257028.75	349762.41	54730.047	246924.35	184405.18	728845.449	1.211336e+05	138256.449	217173.979	381227.82	1079026.71	130516.372	1.950738e+05
FT00206	206445.55	97304.096	17330.480	52770.286	64346.08	712167.72	6003.666	48025.72	841344.15	126954.551	1.821939e+04	10735.200	140713.598	90068.99	10995.90	22025.335	3.700005e+05
FT00210	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
FT00227	613940.43	935076.760	794400.976	2814268.049	2123381.22	1777893.08	107591.662	391781.71	393625.82	760370.085	3.213378e+05	918702.941	652877.439	410643.35	730391.89	290922.043	3.814587e+05
FT00228	221388.37	929136.044	243822.942	21927.820	893010.67	626612.38	55647.235	207208.10	53352.07	475952.822	1.400561e+05	473028.339	297988.354	149251.76	183364.70	118476.159	2.082579e+05
FT00239	61949.97	1756404.369	116173.178	87540.221	151546.97	63225.79	36089.306	63819.06	89499.79	153145.934	5.991865e+04	70555.216	82233.110	26326.56	257623.23	104660.884	5.091533e+04
FT00245	38797.40	38683.180	85420.922	301541.402	152689.60	48787.02	46942.934	543670.33	42579.24	44603.550	6.948973e+04	60052.991	40764.291	423213.30	532491.22	38291.776	9.446117e+04
FT00246	160747.23	257403.732	125539.116	563205.549	513235.73	71727.47	17436.603	194361.76	47129.01	67911.021	4.371184e+04	444789.369	197046.814	81640.28	255775.24	645442.317	3.553732e+04
FT00248	389120.25	51618.815	40008.258	178904.070	135126.39	122217.89	5692.895	26670.19	2130.91	58140.629	2.377895e+04	211381.970	35502.081	27621.44	43530.58	187765.030	3.019649e+04

Taula 2. Matriu filtrada per intensitat i variabilitat biològica>variabilitat analítica; Font: Llenguatge de programació R

Elimino els QC de la matriu filtrada i faig un PCA per visualitzar les mostres filtrades [figura 12]. Observo que tinc *outliers* i reviso la metadata tant de l'urina com de la llista d'aliments que alteren el metaboloma. Les mostres d'orina dels individus 02_11_429, 02_29_238 i 02_09_439 les elimino perquè van menjar xocolata el dia anterior a la recollida de les mostres. Les mostres d'orina dels individus 02_28_399, 01_04_182, 02_12_368, 01_05_155, 01_27_071 les elimino perquè presenten leucòcits en orina. La mostra d'orina de l'individu 02_32_157 l'elimino perquè presenta hemoglobina en sang. La mostra d'orina de l'individu 03_17_045 l'elimino perquè va menjar peix el dia anterior.

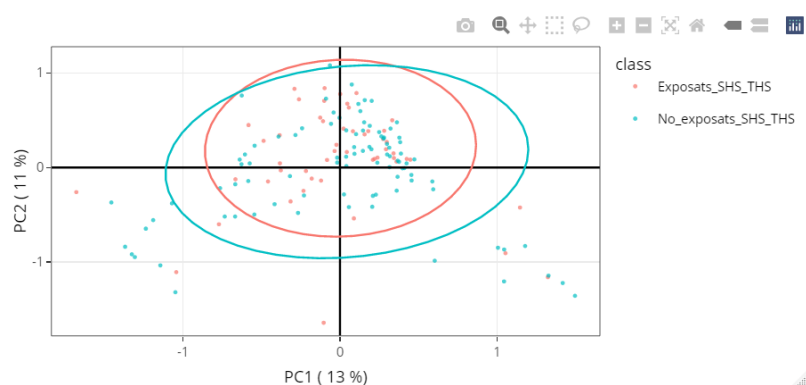


Figura 12. PCA final; Font: Llenguatge de programació R

En el PCA filtrat, podem veure que els grups segueixen sense separar-se entre sí, ja que no només tenim metabòlits exògens, sinó també endògens, que no permeten observar una diferenciació entre grups.

L'últim pas de la filtració consisteix a buscar les masses exactes de les *features* a la HMDB, per tal d'eliminar *features* que no s'hagin trobat en humans. Aquest pas és útil per eliminar *features* provinents de la columna. De les 4530 *features* que tenim a la matriu filtrada, 2032 tenen una massa moisotòpica igual a *features* de la HMDB. El codi de llenguatge R usat es troba en l'annex 11.10.

6.6. Identificació de biomarcadors d'exposició al fum del tabac

Puc generar una matriu de *features*, que em doni la informació sobre la relació massa/càrrega mitjana, temps de retenció, nombre de pics que s'han agrupat per formar la *feature* i presència de la *feature* als grups. El que faig és buscar els biomarcadors de composts del tabac.

Compost	Massa monoisotòpica protonada teòrica	Temps de retenció	de FEATURE
Nicotina-d3	166,1424	281,4 segons	FT03076
Cotinina-d3	180,1216	162,0 segons	FT03802
NNAL-d3	213,1431	110,4 segons	FT05454
Nicotina	163,1235	274,8 segons	FT02934
Cotinina	177,1028	160,2 segons	NA
Hidroxicotinina	193,0977	166,8 segons	FT04395

Taula 3: Massa monoisotòpica protonada teòrica, temps de retenció i feature corresponent dels estàndards i els estàndards interns

Per poder identificar les *features*, la massa protonada i el temps de retenció ha de coincidir entre la *feature* que obtinc a XCMS i els estàndards que he analitzat sota les mateixes condicions cromatogràfiques que les mostres. Busco les masses monoisotòpiques protonades i els temps de retenció a la matriu de *features* per trobar els estàndards interns i els estàndards. Els estàndards interns apareixen a totes les mostres, fet que m'assegura la qualitat analítica del mètode. No hi ha cap *feature* corresponent a la cotinina [Taula 4], això es pel fet que en aquest mètode, la cotinina apareix al principi del cromatograma, on hi ha molt efecte matriu.

	mzmed	mzmin	mzmax	rtmed	rtmin	rtmax	npeaks	Exposats_SHS_THS	No_exposats_SHS_THS	QC
FT03656	177.0909	177.0907	177.0912	125.71265	124.07108	126.94678	83	25	34	24
FT03657	177.0950	177.0910	177.0953	323.44019	321.44220	343.51285	118	29	50	23
FT03658	177.1054	177.1019	177.1056	327.36699	313.35626	364.11633	92	22	33	22
FT03659	177.1062	177.1059	177.1064	376.72353	370.74280	383.66718	166	49	91	24
FT03660	177.1121	177.1118	177.1123	129.00101	126.86745	130.29137	162	49	89	24
FT03661	177.1121	177.1120	177.1123	169.04837	167.79666	170.78857	133	39	70	24

Taula 4. No hi ha cap *feature* corresponent a la cotinina; Font: Llenguatge de programació R

Hem pogut identificar la nicotina i la hidroxicotinina [Taula 5 i 6 respectivament]. La presència de la nicotina és ubiqüa, està present en més del 80% tant dels nens exposats com no exposats, fet que recolza que l'exposició passiva al fum del tabac està molt present en la població pediàtrica. També hem identificat la hidroxicotinina, un dels metabòlits de la cotinina en un 65% dels nens exposats i un 37% en els no exposats al fum del tabac.

	mzmed	mzmin	mzmax	rtmed	rtmin	rtmax	npeaks	Exposats_SHS_THS	No_exposats_SHS_THS	QC
FT02934	163.1231	163.1228	163.1233	277.69711	259.75342	293.38559	151	43	79	24

Taula 5. *Feature* de la Nicotina; Font: Llenguatge de programació R

	mzmed	mzmin	mzmax	rtmed	rtmin	rtmax	npeaks	Exposats_SHS_THS	No_exposats_SHS_THS	QC
FT04395	193.0971	193.0967	193.0973	166.00789	152.12521	188.87152	117	33	35	24

Taula 6. *Feature* de l'hidroxicotina; Font: Llenguatge de programació R

Per tal d'observar visualment si hi ha diferències pel que fa a la intensitat de la nicotina i la hidroxicotina entre els nens exposats i no exposats al fum del tabac, faig un *boxplot* amb R (El codi de llenguatge R utilitzat es troba a l'Annex 11.11). Podem veure que hi ha diferències i que els nens exposats presenten tant nicotina com hidroxicotina en una intensitat superior envers els nens no exposats [figures 13 i 14 respectivament].

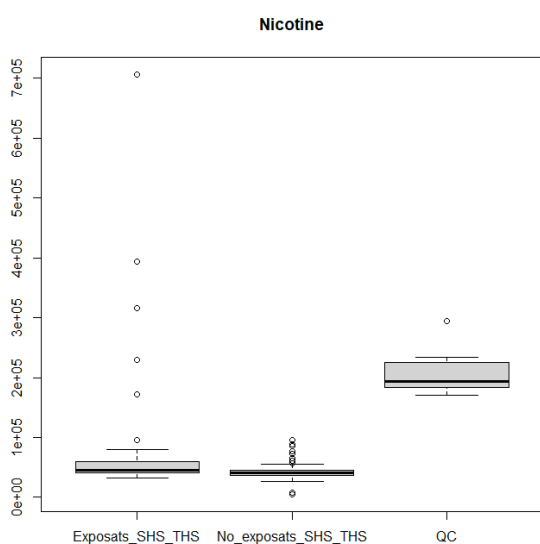


Figura 13: *Boxplot* Nicotina; Font: Llenguatge de programació R

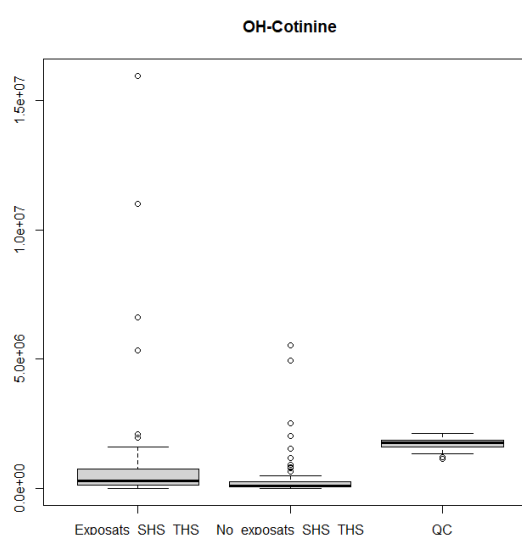


Figura 14: *Boxplot* hidroxicotina; Font: Llenguatge de programació R

7. CONCLUSIONS

La metabolòmica no dirigida és una eina potent per a l'exploració de les possibles alteracions metabòliques en orina de nens i nenes associades a l'exposició a contaminants ambientals. El *software* HERMES és una eina excepcional per seleccionar els ions per obtenir els seus espectres de MS².

XCMS proporciona eines per poder resoldre les desviacions en el temps de retenció pròpies de l'anàlisi i a més permet agrupar aquests pics que presenten certes desalineacions entre sí per formar les *features*.

En aquest estudi hem observat que la intensitat relativa de la nicotina és superior en nens exposats que en no exposats al fum del tabac. Tot i això, hem confirmat l'exposició ubicua nicotina, ja que es troba en més del 80% tant de nens exposats com no exposats al fum del tabac, fet que recolza que l'exposició al fum del tabac està present en la població més vulnerable. D'altra banda, hem identificat hidroxicotinina tant en nens exposats com no exposats, tot i que la proporció d'aquesta era superior en nens exposats al fum del tabac.

8. PERSPECTIVES DE FUTUR

En aquest treball presento l'inici del processat de dades de les orines d'aquest cohort. Per tal de poder avaluar les possibles alteracions metabòliques en aquests nens i nenes, s'hauran de realitzar les següents accions:

1. Anàlisi estadístic univariant i multivariant de les dades de metabolòmica no dirigida per tal de seleccionar les *features* rellevants amb les dades de la cohort, fent especial incidència amb la convivència amb fumadors, classe social i nivell d'estudis dels progenitors.
2. Identificació de les *features* rellevants utilitzant els espectres de MS² seleccionats per HERMES. La identificació es farà comparant els espectres adquirits amb espectres experimentals de bases de dades (com ara METLIN o NIST) o mitjançant l'anàlisi d'estàndards dels metabòlits.
3. Estudi de les vies metabòliques alterades per l'exposició.

9. BIBLIOGRAFIA

1. Torres, S., Merino, C., Paton, B., Correig, X., & Ramírez, N. (2018). Biomarkers of exposure to secondhand and thirdhand Tobacco smoke: Recent advances and future perspectives. In *International Journal of Environmental Research and Public Health* (Vol. 15, Issue 12). MDPI AG. <https://doi.org/10.3390/ijerph15122693>
2. Jacob, P., Benowitz, N. L., Destailats, H., Gundel, L., Hang, B., Martins-Green, M., Matt, G. E., Quintana, P. J. E., Samet, J. M., Schick, S. F., Talbot, P., Aquilina, N. J., Hovell, M. F., Mao, J. H., & Whitehead, T. P. (2017). Thirdhand smoke: New evidence, challenges, and future directions. *Chemical Research in Toxicology*, 30(1), 270–294. <https://doi.org/10.1021/acs.chemrestox.6b00343>

3. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans., World Health Organization., & International Agency for Research on Cancer. (2004). *Tobacco smoke and involuntary smoking*. IARC Press.
4. McGrath, J. J., Racicot, S., Okoli, C. T. C., Hammond, S. K., & O'loughlin, J. (2018). Airborne Nicotine, Secondhand Smoke, and Precursors to Adolescent Smoking. In *PEDIATRICS* (Vol. 141, Issue S1). http://publications.aap.org/pediatrics/article-pdf/141/Supplement_1/S63/903459/peds_20171026j.pdf
5. Adhami, N., Chen, Y., & Martins-Green, M. (2017). Biomarkers of disease can be detected in mice as early as 4 weeks after initiation of exposure to third-hand smoke levels equivalent to those found in homes of smokers. *Clinical Science*, 131(19), 2409–2426. <https://doi.org/10.1042/CS20171053>
6. Moon, K. A., Rule, A. M., Magid, H. S., Ferguson, J. M., Susan, J., Sun, Z., Torrey, C., Abubaker, S., Levshin, V., Çarkoglu, A., Radwan, G. N., El-Rabbat, M., Cohen, J. E., Strickland, P., Breysse, P. N., & Navas-Acien, A. (2018). Biomarkers of secondhand smoke exposure in waterpipe tobacco venue employees in Istanbul, Moscow, and Cairo. *Nicotine and Tobacco Research*, 20(4), 482–491. <https://doi.org/10.1093/ntr/ntx125>
7. Lletjós, P., Continente, X., Arechavala, T., Fernández, E., Schiaffino, A., Pérez-Ríos, M., & López, M. J. (2020). Association between exposure to second-hand smoke and health status in children. *Gaceta Sanitaria*, 34(4), 363–369. <https://doi.org/10.1016/j.gaceta.2018.10.006>
8. Repace, J. (2006). Exposure to Secondhand Smoke. In *Exposure Analysis* (pp. 201–235). CRC Press. <https://doi.org/10.1201/9781420012637.ch9>
9. Com, I. (2018). *AIR POLLUTION AND CHILD HEALTH*. <http://apps.who.int/bookorders>.
10. Ambient air pollution: A global assessment of exposure and burden of disease. (n.d.).
11. Mahabee-Gittens, E. M., Merianos, A. L., & Matt, G. E. (2018). Preliminary evidence that high levels of nicotine on children's hands may contribute to overall

tobacco smoke exposure. *Tobacco Control*, 27(2), 217–219.
<https://doi.org/10.1136/tobaccocontrol-2016-053602>

12. Dunn, W. B., Broadhurst, D., Begley, P., Zelena, E., Francis-Mcintyre, S., Anderson, N., Brown, M., Knowles, J. D., Halsall, A., Haselden, J. N., Nicholls, A. W., Wilson, I. D., Kell, D. B., & Goodacre, R. (2011). Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature Protocols*, 6(7), 1060–1083.
<https://doi.org/10.1038/nprot.2011.335>

13. Yanes, O., Tautenhahn, R., Patti, G. J., & Siuzdak, G. (2011). Expanding coverage of the metabolome for global metabolite profiling. *Analytical Chemistry*, 83(6), 2152–2161. <https://doi.org/10.1021/ac102981k>

14. Capriotti, A. L., Cannazza, G., Catani, M., Cavaliere, C., Cavazzini, A., Cerrato, A., Citti, C., Felletti, S., Montone, C. M., Piovesana, S., & Laganà, A. (2021). Recent applications of mass spectrometry for the characterization of cannabis and hemp phytocannabinoids: From targeted to untargeted analysis. In *Journal of Chromatography A* (Vol. 1655). Elsevier B.V.
<https://doi.org/10.1016/j.chroma.2021.462492>

15. Arija, V., Fargas, F., March, G., Abajo, S., Basora, J., Canals, J., Ribot, B., Aparicio, E., Serrat, N., Hernández-Martínez, C., & Aranda, N. (2014). *Adapting iron dose supplementation in pregnancy for greater effectiveness on mother and child health: protocol of the ECLIPSES randomized clinical trial*.
<http://www.biomedcentral.com/1471-2393/14/33>

16. Southam, A. D., Haglington, L. D., Najdekr, L., Jankevics, A., Weber, R. J. M., & Dunn, W. B. (2020). Assessment of human plasma and urine sample preparation for reproducible and high-throughput UHPLC-MS clinical metabolic phenotyping. *Analyst*, 145(20), 6511–6523. <https://doi.org/10.1039/d0an01319f>

17. Giné, R., Capellades, J., Badia, J. M., Vughs, D., Schwaiger-Haber, M., Vinaixa, M., Brunner, A. M., Patti, G. J., & Yanes, O. (n.d.). *HERMES: a molecular formula-oriented method to target the metabolome*.
<https://doi.org/10.1101/2021.03.08.434466>

10. AUTOAVALUACIÓ

La meua estància al grup MIL@b ha estat molt enriquidora, he format part d'un grup de recerca multidisciplinari que m'ha donat el seu suport i brindat coneixements en l'àmbit científic. Des del primer moment m'he sentit una més del grup de recerca i he pogut comptar sempre amb el recolzament de la meua tutora professional Noelia Ramírez González i altres estudiants com la Carla Merino o la Mireia Palau.

He enriquit els meus coneixements tant en tasques de laboratori, amb l'extracció dels metabòlits de les mostres d'orina i l'ús de l'equip LC-ESI-MS, com en el processament de les dades obtingudes mitjançant XCMS i HERMES. He reforçat el meu criteri analític i estic molt orgullosa de l'aprenentatge que he assolit durant el transcurs del treball de final de grau.

11. ANNEXOS

11.01. Metadades exposició al fum del tabac

ID ECLIPSES	Convivència fumadors	Fumar pare	Pare fuma davant del nen/a	Nº cigarretes pare	Pare fuma a casa	Fumar mare	Mare fuma davant del nen/a	Nº cigarretes mare	Mare fuma a casa
01-02-069	Sí	Sí	Sí	5,00	Sí
01-03-057	Sí	Sí	No	5,00	No
01-02-027	No
01-02-062	No
01-03-136	Sí	Sí	No	15,00	Sí
01-03-169	Sí	.	Sí	2,00	Sí
01-02-140	No
01-03-174	Sí	Sí	12,00	.
01-03-204	Sí	.	Sí	7,50	-	.	Sí	5,50	-
01-03-144	No
01-03-159	No
01-04-030	No
01-03-208	Sí	.	Sí	2,00	No	.	Sí	8,00	No
01-04-146	Sí	Sí	No	20,00	Sí	Sí	No	5,00	Sí
01-04-089	No
01-04-182	No
01-05-072	Sí
01-04-197	No
01-05-100	No
01-05-073	Sí	Sí	No	10,00	Sí
01-06-064	Sí	Sí	No	15,00	Sí	Sí	.	.	.
01-05-155	No
01-06-078	No
01-06-096	No
01-06-097	No
02-09-057	Sí	Sí	No	1,00	Sí
01-06-161	No
02-09-117	Sí	Sí	No	10,00
01-18-039	No

01-18-059	No
01-18-099	No
02-09-329	Sí	.	Sí	.	Sí
01-18-112	No
01-18-179	No
02-09-382	Sí	Sí	Sí	30,00	Sí
02-09-456	Sí	Sí	No	1,00	No
01-26-054	No
01-27-071	No
02-12-184	Sí	Sí	No	10,00	No
01-27-077	No
02-13-048	Sí	Sí	No	4,00	No	Sí	No	6,00	No
01-38-001	No
02-13-471	Sí
02-28-043	Sí
02-09-239	No
02-09-298	No
02-09-332	No
02-09-335	No
02-28-096	Sí
02-09-364	No
02-28-446	Sí	Sí	No	11,00	No
02-09-439	No
02-29-238	Sí	Sí	No	5,00
02-10-365	No
02-29-418	Sí
02-11-158	No
02-11-177	No
02-11-203	No
02-11-295	No
02-11-429	No
02-11-436	No
02-11-444	No
02-12-122	No
02-30-110	Sí
02-30-163	Sí	Sí	Sí	10,00
02-12-368	No
02-12-413	No
02-12-432	No
02-13-330	No
02-30-208	Sí
02-13-334	No
02-32-157	Sí	Sí	Sí	5,00
02-32-340	Sí	Sí	No	20,00	Sí	.	No	30,00	Sí
02-28-367	No
02-28-399	No
02-29-224	No
02-37-269	Sí	Sí	.	.	Sí	Sí	.	17,50	Sí
02-29-305	No
02-29-372	No
02-43-395	Sí	Sí	No	8,00	No	Sí	No	8,00	1
02-43-397	Sí	Sí	Sí	10,00	No
02-30-085	No
02-30-162	No
02-43-398	Sí	Sí	Sí	8,50	Sí

02-30-216	No
02-30-228	No
03-17-020	Sí	Sí	Sí	40,00
04-18-002	Sí	Sí	.	20,00	Sí
02-32-231	No
02-32-234	No
04-18-009	Sí	Sí	No	6,00	Sí
02-32-250	No
02-37-147	No
03-17-003	No
04-18-011	Sí	Sí	No	20,00	Sí	Sí	No	20,00	Sí
04-18-016	Sí	.	.	15,00	.	Sí	.	10,00	.
05-20-006	Sí	Sí	Sí	20,00	Sí
03-17-026	No
03-17-045	No
03-17-052	No
03-17-055	No
05-20-030	Sí	Sí	No	3,00	.
06-21-017	Sí	Sí	Sí	7,00	.
03-17-058	No
04-18-004	No
06-22-003	Sí	Sí	No	10,00	.	Sí	No	15,00	.
04-18-012	No
04-18-029	No
04-18-048	No
04-18-049	No
04-18-051	No
04-18-054	No
05-19-038	No
05-20-005	No
05-20-009	No
05-20-017	No
05-20-020	No
05-20-036	No
06-21-005	No
06-21-007	No
06-22-010	No
07-23-007	No
07-23-011	No
07-23-012	No
09-40-007	No
06-22-004	Sí	Sí	No	12,00	No	Sí	No	5,00	No
06-22-006	Sí	Sí	Sí	10,00	Sí	Sí	Sí	10,00	Sí
09-40-019	No
10-38-003	No
10-38-017	No
06-22-022	Sí	.	No	15,00	Sí	.	No	15,00	Sí
07-23-002	Sí	Sí	.	15,00	.	Sí	Sí	5,00	.
10-38-029	No
10-38-030	No
11-39-006	No
07-23-021	Sí	Sí	Sí	20,00	Sí
11-39-015	No
07-33-004	Sí	Sí	No	10,00	No
09-40-011	Sí	Sí	Sí	15,00	.	Sí	Sí	12,00	.

13-44-001	No
10-38-028	Sí	Sí	No	25,00	No
13-45-023	Sí	Sí	No	10,00	.
13-44-014	No
13-46-006	No
13-46-011	No
13-46-025	No

Taula 7: Metadades sobre la exposició dels nens i nenes al fum passiu del tabac, aquesta taula inclou la informació sobre el membra familiar fumador, el nombre de cigarretes que es fuma al dia, si fuma a casa i si ho fa davant del menor. Font: pròpia

11.02. Metadades ingesta aliments que alteren el metaboloma

ID ECLIPSES	Peix	Xocolata	Begudes amb gas
05-19-038	No consum	No consum	No consum
13-46-025	No consum	No consum	No consum
10-38-028	No consum	No consum	No consum
02-13-471	No consum	No consum	No consum
11-39-015	No consum	Consum	No consum
03-17-045	Consum	No consum	No consum
02-09-439	No consum	Consum	No consum
09-40-011	No consum	Consum	No consum
09-40-012	No consum	Consum	No consum
02-29-418	No consum	No consum	No consum
01-06-161	No consum	No consum	No consum
01-05-100	No consum	No consum	No consum
01-03-136	No consum	No consum	No consum
01-18-099	No consum	No consum	No consum
01-02-140	No consum	No consum	No consum
10-38-017	No consum	Consum	No consum
01-18-112	No consum	No consum	No consum
01-02-157	No consum	No consum	No consum
12-42-001	No consum	Consum	No consum
13-44-014	No consum	No consum	No consum
02-28-446	No consum	No consum	No consum
02-29-305	No consum	No consum	No consum
02-43-397	Consum	No consum	No consum
02-11-295	No consum	No consum	No consum
01-18-179	No consum	No consum	No consum
06-22-022	No consum	No consum	No consum
02-11-436	No consum	Consum	No consum
02-09-364	No consum	No consum	No consum
02-09-329	No consum	No consum	No consum
02-28-399	No consum	No consum	No consum
02-10-365	No consum	Consum	No consum
02-11-429	No consum	Consum	No consum
06-21-017	No consum	No consum	No consum
02-09-456	No consum	No consum	No consum

ID ECLIPSES	Peix	Xocolata	Begudes amb gas
01-05-155	No consum	No consum	No consum
01-03-159	Consum	No consum	No consum
02-09-332	No consum	Consum	No consum
02-13-334	No consum	Consum	No consum
02-13-330	No consum	No consum	No consum
01-05-155	No consum	No consum	No consum
01-03-159	Consum	No consum	No consum
02-09-332	No consum	Consum	No consum
06-22-010	No consum	No consum	No consum
02-32-234	No consum	No consum	No consum
02-09-239	No consum	Consum	No consum
04-18-009	No consum	No consum	No consum
02-30-228	No consum	No consum	No consum
02-09-244	No consum	No consum	No consum
07-33-004	Consum	Consum	No consum
07-23-021	No consum	No consum	No consum
07-23-012	No consum	No consum	No consum
01-05-073	Consum	No consum	No consum
03-17-026	No consum	No consum	No consum
02-32-231	No consum	Consum	No consum
01-18-039	No consum	No consum	No consum
01-05-072	No consum	No consum	No consum
02-12-145	No consum	No consum	No consum
04-18-012	No consum	No consum	No consum
02-11-177	No consum	No consum	No consum
09-40-007	No consum	No consum	No consum
01-27-077	No consum	No consum	No consum
04-18-016	No consum	No consum	No consum
02-11-203	Consum	No consum	No consum
01-04-089	No consum	No consum	No consum
06-22-003	No consum	No consum	No consum
02-28-043	No consum	No consum	No consum
01-02-062	No consum	Consum	No consum
10-38-003	No consum	No consum	No consum

02-29-372	No consum	No consum	No consum
09-40-019	No consum	No consum	No consum
02-12-432	No consum	No consum	No consum
02-11-444	No consum	Consum	No consum
02-12-368	No consum	No consum	No consum
02-28-367	No consum	No consum	No consum
02-28-287	No consum	Consum	No consum
02-12-413	No consum	No consum	No consum
02-32-340	No consum	No consum	No consum
02-09-298	No consum	No consum	No consum
02-13-334	No consum	Consum	No consum
02-13-330	No consum	No consum	No consum

05-20-009	No consum	No consum	No consum
02-30-110	No consum	No consum	No consum
01-26-054	No consum	No consum	No consum
01-02-069	No consum	No consum	No consum
02-09-223	No consum	No consum	No consum
02-09-057	No consum	No consum	No consum
05-20-017	No consum	No consum	No consum
05-20-006	No consum	No consum	No consum
01-06-097	No consum	No consum	No consum
01-03-057	Consum	No consum	No consum
01-06-078	No consum	No consum	No consum
04-18-011	No consum	Consum	No consum

Taula 8. Metadades sobre la ingesta el dia anterior d'aliments que alteren el metaboloma (peix, xocolata i begudes amb gas); Font: pròpia

11.03. Metadades urianàlisi

ID ECLIPSES	SANG/HEMOGLOBINA	DENSITAT RELATIVA	LEUCÒCITS
01-02-027	-	1.015	-
01-02-062	-	1.025	-
01-02-069	-	1.025	-
01-02-140	-	1.015	-
01-03-057	-	1.030	-
01-03-136	-	1.025	-
01-03-144	-	1.020	-
01-03-159	-	1.020	-
01-03-169	-	1.025	-
01-03-174	-	1.025	-
01-03-204	-	1.030	25
01-03-208	-	1.010	-
01-04-030	-	1.020	-
01-04-089	-	1.010	-
01-04-146	-	1.030	-
01-04-182	-	1.020	(++) 75 WBC/μl
01-04-197	-	1.025	-
01-05-072	-	1.025	-
01-05-073	-	1.030	-
01-05-100	-	1.025	-
01-05-155	-	1.025	(++) 75 WBC/μl
01-06-064	-	1.020	-
01-06-078	-	1.025	-
01-06-096	-	1.015	-
01-06-097	-	1.025	-
01-06-161	-	1.030	-
01-18-039	-	1.020	-
01-18-059	-	1.015	-
01-18-099	-	1.030	-

ID ECLIPSES	SANG/HEMOGLOBINA	DENSITAT RELATIVA	LEUCÒCITS
02-30-085	-	1.030	-
02-30-110	-	1.025	(++) 25 WBC/μl
02-30-162	-	1.020	-
02-30-163	-	1.025	(++) 75Leu/μl
02-30-208	-	1.025	-
02-30-216	-	1.030	-
02-30-228	-	1.025	-
02-32-157	Hb +3 (Ca. 50 Ery/μl)	1.025	-
02-32-231	-	1.030	-
02-32-234	-	1.030	-
02-32-250	-	1.025	(+) 10-25 Leu/μl
02-32-340	-	1.025	-
02-37-147	-	1.015	-
02-37-269	-	1.030	-
02-43-395	-	1.025	-
02-43-397	-	1.030	-
02-43-398	-	1.025	-
03-17-003	-	1.030	-
03-17-020	-	1.030	-
03-17-026	-	1.030	-
03-17-045	-	1.020	-
03-17-052	-	1.030	-
03-17-055	-	1.030	-
03-17-058	-	1.010	-
04-18-002	-	1.020	-
04-18-004	-	1.015	(+) 10-25 Leu/μl
04-18-009	-	1.030	-
04-18-011	-	1.025	-
04-18-012	-	1.005	-

01-18-112	-	1.030	-	04-18-016	-	1.025	-
01-18-179	-	1.025	-	04-18-029	-	1.030	-
01-26-054	-	1.020	-	04-18-048	-	1.025	-
01-27-071	-	1.015	(+++) 500 WBC/ μ l	04-18-049	-	1.015	-
01-27-077	-	1.015	-	04-18-051	-	1.025	-
01-38-001	-	1.015	-	04-18-054	-	1.030	-
02-09-057	-	1.020	-	05-19-038	-	1.020	-
02-09-117	-	1.025	-	05-20-005	-	1.020	-
02-09-239	-	1.030	-	05-20-006	-	1.020	(+++) 500 Leu/ μ l
02-09-298	-	1.015	-	05-20-030	-	1.030	-
02-09-329	Hb +1 (Ca. 10 Ery/ μ l)	1.020	-	05-20-009	-	1.030	-
02-09-332	-	1.030	-	05-20-017	Hb +1 (Ca. 10 Ery/ μ l)	1.030	-
02-09-335	-	1.025	-	05-20-020	-	1.030	-
02-09-364	-	1.030	-	05-20-036	-	1.030	-
02-09-382	-	1.025	-	06-21-005	-	1.030	-
02-09-439	-	1.025	-	06-21-007	-	1.025	(++) 75 Leu/ μ l
02-09-456	-	1.025	-	06-22-003	-	1.030	-
02-10-365	-	1.030	-	06-22-004	-	1.030	-
02-11-158	-	1.025	-	06-22-006	-	1.010	-
02-11-177	-	1.030	-	06-22-010	-	1.030	-
02-11-203	-	1.025	-	06-22-022	-	1.025	-
02-11-295	-	1.020	(+) 10-25 Leu/ μ l	07-23-002	-	1.025	-
02-11-429	-	1.030	-	07-23-007	-	1.030	-
02-11-436	-	1.025	-	07-23-011	-	1.025	-
02-11-444	-	1.025	-	07-23-012	-	1.015	-
02-12-122	-	1.015	-	07-23-021	-	1.020	-
02-12-184	-	1.025	-	07-33-004	-	1.030	-
02-12-368	-	1.025	(+) 10-25 Leu/ μ l	09-40-007	-	1.030	-
02-12-413	-	1.015	-	09-40-011	-	1.025	-
02-12-432	-	1.025	-	09-40-019	-	1.025	-
02-13-048	-	1.025	-	10-38-003	-	1.030	-
02-13-330	-	1.025	(+) 10-25 Leu/ μ l	10-38-017	-	1.025	(++) 75 WBC/ μ l
02-13-334	-	1.010	-	10-38-028	(++) 50 RBC/ μ l	1.025	(++) 75 WBC/ μ l
02-13-471	-	1.020	-	10-38-029	-	1.030	(++) 75 WBC/ μ l
02-28-043	-	1.025	-	10-38-030	-	1.020	(+++) 500 Leu/ μ l
02-28-096	-	1.020	-	11-39-006	-	1.030	-
02-28-367	-	1.025	-	11-39-015	-	1.030	-
02-28-399	-	1.025	(+) 10-25 Leu/ μ l	12-42-001	-	1.020	-
02-28-446	-	1.025	-	13-44-014	-	1.020	(++) 75 WBC/ μ l
02-29-224	-	1.025	-	13-45-023	-	1.025	-
02-29-238	-	1.030	-	13-46-006	-	1.025	-
02-29-305	-	1.025	-	13-46-011	-	1.025	-
02-29-372	-	1.030	-	13-46-025	-	1.030	-
02-29-418	-	1.030	-	13-46-025	-	1.030	-

Taula 9: Metadades de l'urinalísi de les mostres d'orina, aquest inclou la presència d'hemoglobina en sang, la densitat relativa i la presència de leucòcits en orina. Font: pròpia

11.04. Integració de pics i %RSD de la nicotina-d3

NOM DEL COMPOST		
Nicotine-d3	Average Response Factor = 4.01616e+007 %RSD = 19.2	
Filename	Area	RT
01_02_027	210300699	4,77
01_02_062	207713011	4,75
01_02_069	171632100	4,78
01_02_140	196999384	4,73
01_03_057	194663775	4,75
01_03_136	214038716	4,74
01_03_144	221901549	4,74
01_03_159	226092212	4,74
01_03_169	154681912	4,77
01_03_174	171872517	4,77
01_03_204	132407354	4,78
01_03_208	227272449	4,73
01_04_030	200639132	4,75
01_04_089	212575623	4,76
01_04_146	237342387	4,74
01_04_182	187013177	4,77
01_04_197	161880908	4,77
01_05_072	209019703	4,76
01_05_073	215518193	4,74
01_05_100	205955999	4,74
01_05_155	225627088	4,73
01_06_064	217815982	4,75
01_06_078	202016117	4,75
01_06_096	218530396	4,74
01_06_097	215590886	4,75
01_06_161	215532180	4,74
01_18_039	192017609	4,74
01_18_059	216297244	4,75
01_18_099	214383204	4,75
01_18_112	228160955	4,74
01_18_179	232699079	4,74
01_26_054	195020377	4,77
01_27_071	223716984	4,74
01_27_077	201489054	4,77
01_38_001	198915761	4,75
02_09_057	208026565	4,76
02_09_117	200136426	4,75
02_09_239	12877230	4,76
02_09_298	217911120	4,73
02_09_329	228231586	4,74
02_09_332	227318964	4,73
02_09_335	220908341	4,73
02_09_364	235239072	4,75
02_09_382	238058601	4,72
02_09_439	216964746	4,73
02_09_456	244519705	4,73
02_10_365	229988078	4,75
02_11_158	195543510	4,77
02_11_177	193300659	4,77
02_11_203	170496364	4,77
02_11_203	170496364	4,77
02_11_295	243029699	4,73
02_11_429	222425921	4,73
02_11_436	222780569	4,74
02_11_444	235113854	4,73
02_12_122	199529856	4,75

02_12_184	204638976	4,76
02_12_368	217577799	4,74
02_12_413	231158185	4,74
02_12_432	227654285	4,74
02_13_048	184245148	4,76
02_13_048	184245148	4,76
02_13_330	231154093	4,73
02_13_334	237017305	4,72
02_13_471	237914478	4,73
02_28_043	213020448	4,75
02_28_096	176134831	4,77
02_28_367	243770122	4,74
02_28_399	226881779	4,73
02_28_446	210560822	4,74
02_29_224	216684375	4,75
02_29_238	204222770	4,74
02_29_244	214269954	4,77
02_29_305	241757058	4,74
02_29_372	226775723	4,74
02_29_418	221813487	4,74
02_30_085	208439266	4,76
02_30_110	154676990	4,78
02_30_162	201550195	4,76
02_30_163	187708625	4,77
02_30_208	183201812	4,76
02_30_216	200803797	4,75
02_30_228	196545950	4,75
02_32_157	213779562	4,76
02_32_227	210872482	4,76
02_32_231	202142699	4,75
02_32_234	215128508	4,75
02_32_250	225591941	4,74
02_32_340	239465404	4,72
02_37_147	211184333	4,75
02_37_269	221016830	4,74
02_43_395	232070473	4,73
02_43_397	230972327	4,73
02_43_398	233935749	4,72
03_17_003	191424	5,10
03_17_020	204994769	4,76
03_17_026	211951965	4,74
03_17_045	174196100	4,77
03_17_052	116688551	4,77
03_17_055	215216	4,94
03_17_058	183188985	4,76
03_36_006	197858253	4,75
04_18_002	184024358	4,76
04_18_004	213237306	4,75
04_18_009	204578357	4,75
04_18_011	231126113	4,74
04_18_012	114090	5,10
04_18_016	205368939	4,76
04_18_029	206007387	4,74
04_18_048	189450816	4,77
04_18_049	183906530	4,78
04_18_051	181585993	4,77
04_18_054	169851274	4,78
05_19_038	228812732	4,77
05_20_005	215213886	4,75
05_20_006	215625389	4,75
05_20_009	224858067	4,75
05_20_017	194903934	4,76
05_20_020	213977077	4,75
05_20_023	177164493	4,77
05_20_030	199289660	4,74

05_20_036	164321670	4,77
06_21_005	217626559	4,74
06_21_007	200243849	4,75
06_21_017	239295595	4,74
06_22_003	195170969	4,78
06_22_004	187119973	4,77
06_22_006	209923652	4,76
06_22_010	186956762	4,75
06_22_022	216225583	4,74
07_23_002	202062027	4,76
07_23_007	197576449	4,74
07_23_011	185358801	4,78
07_23_012	203072715	4,75
07_23_021	188643198	4,76
07_33_004	212863023	4,75
09_40_007	187884424	4,77
09_40_011	218445578	4,74
09_40_012	210043613	4,74
09_40_019	224409304	4,73
10_38_003	189501105	4,76
10_38_017	220266943	4,74
10_38_028	194573293	4,77
10_38_029	160228870	4,77
10_38_030	180992527	4,78
11_39_006	201382136	4,75
11_39_015	141210676	4,77
12_42_001	217917175	4,74
12_56_015	192692303	4,77
13_44_001	205188788	4,75
13_44_014	181945571	4,78
13_45_023	178674782	4,77
13_46_006_b	174610609	4,77
13_46_011_b	165431272	4,78
13_46_025	137069776	4,77

Taula 10: Integració dels pic de nicotina-d3, temps de retenció dels pics àrees de les integracions; Font: pròpia

11.05. Integració de pics i %RSD de la cotinina-d3

NOM DEL COMPOST		
Average Response Factor = 4.78671e+008		
Cotinina-d3	%RSD = 17.6	
Filename	Area	RT
01_02_027	2796724598	2,70
01_02_062	2179604828	2,69
01_02_069	2709812762	2,72
01_02_140	2301581500	2,69
01_02_157	2590761820	2,69
01_03_057	2425778659	2,70
01_03_136	2712480219	2,69
01_03_144	2516608700	2,68
01_03_159	2561910644	2,70
01_03_169	2152130805	2,71
01_03_174	2186950093	2,71
01_03_204	2220168463	2,73
01_03_208	2921698069	2,68
01_04_030	2431121298	2,70
01_04_089	2939275892	2,69
01_04_146	2562112465	2,68
01_04_182	2636423936	2,70
01_04_197	2354361704	2,71

01_05_072	2387025047	2,70
01_05_073	2243680703	2,67
01_05_100	2534365070	2,69
01_05_155	2207795151	2,69
01_06_064	2530030549	2,69
01_06_078	2563095239	2,69
01_06_096	2607515106	2,69
01_06_097	2555048625	2,69
01_06_161	2439300830	2,69
01_18_039	2018192769	2,68
01_18_059	2706863303	2,69
01_18_099	2428282368	2,69
01_18_112	2333108029	2,69
01_18_179	2421226162	2,68
01_26_054	2808231928	2,70
01_27_071	2605272719	2,68
01_27_077	2383586674	2,69
01_38_001	2792358474	2,70
02_08_093	2131003170	2,69
02_09_057	2396472219	2,69
02_09_100	2281736938	2,71
02_09_117	2663196426	2,70
02_09_165	2489596333	2,70
02_09_223	2498629867	2,70
02_09_239	258592161	2,72
02_09_298	2510842226	2,68
02_09_329	2646710551	2,67
02_09_332	2499900343	2,67
02_09_335	2264712788	2,67
02_09_345	2558278892	2,68
02_09_364	2550893160	2,68
02_09_382	2385195765	2,67
02_09_439	2278780025	2,68
02_09_456	2386390364	2,69
02_10_365	2405613126	2,69
02_11_158	2704507738	2,70
02_11_177	2637983762	2,69
02_11_203	2167283113	2,69
02_11_203	2167283113	2,69
02_11_230	2314266985	2,69
02_11_295	2471779929	2,68
02_11_429	2445381995	2,68
02_11_436	2199348915	2,68
02_11_444	2462014870	2,68
02_12_076	2497269451	2,69
02_12_122	2569221094	2,70
02_12_133	2556719037	2,70
02_12_145	2639019921	2,71
02_12_184	2859797999	2,71
02_12_368	2516639588	2,68
02_12_413	2709114359	2,68
02_12_432	2286337180	2,68
02_13_048	2357033982	2,69
02_13_048	2402244808	2,69
02_13_330	2749733380	2,68
02_13_334	2473979103	2,68
02_13_471	2575732024	2,68
02_28_043	2627293894	2,69
02_28_063	2456071228	2,71
02_28_096	2529345615	2,71
02_28_109	2506728152	2,69
02_28_249	2269360112	2,70
02_28_287	1980036792	2,68
02_28_367	2494336691	2,69
02_28_399	2202866905	2,69

02_28_446	2284879075	2,68
02_29_224	2494998884	2,70
02_29_238	2287336420	2,69
02_29_244	1766253243	2,69
02_29_305	2607383920	2,69
02_29_372	2554338190	2,68
02_29_418	2252075958	2,67
02_30_085	2562378373	2,70
02_30_110	2511912889	2,70
02_30_162	2388238865	2,71
02_30_163	2338998083	2,71
02_30_208	2432694129	2,71
02_30_216	2388494059	2,68
02_30_228	2152618383	2,68
02_32_157	2650102278	2,70
02_32_227	2753317021	2,68
02_32_231	2121582192	2,69
02_32_234	2541183194	2,68
02_32_250	2351098733	2,69
02_32_340	2033601258	2,68
02_37_147	2407037665	2,69
02_37_269	2453319036	2,67
02_43_395	2285160049	2,69
02_43_397	2447555053	2,68
02_43_398	2536065038	2,68
03_17_003	306778	2,75
03_17_020	2358545079	2,69
03_17_026	2240574014	2,69
03_17_045	2288887038	2,72
03_17_052	2456718142	2,71
03_17_055	638285	2,74
03_17_058	2445070003	2,70
03_36_006	2905803696	2,70
04_18_002	2410814210	2,69
04_18_004	2614475899	2,70
04_18_009	2211525873	2,67
04_18_011	2554828945	2,69
04_18_012	393611	2,74
04_18_016	2409418035	2,71
04_18_029	2428621421	2,67
04_18_048	2671325890	2,70
04_18_049	2633529178	2,70
04_18_051	2554329449	2,72
04_18_054	2559702210	2,71
05_19_038	2462896704	2,72
05_20_005	2618571590	2,69
05_20_006	2766117868	2,70
05_20_009	2697034600	2,68
05_20_017	2407389554	2,70
05_20_020	2233740483	2,68
05_20_023	2234779022	2,70
05_20_030	2220628324	2,69
05_20_036	2160576967	2,70
06_21_005	2711871627	2,71
06_21_007	2638705702	2,69
06_21_017	2426555250	2,68
06_22_003	2464934762	2,70
06_22_004	2487919613	2,71
06_22_006	2504246995	2,70
06_22_010	2076396815	2,69
06_22_022	1907652288	2,67
07_23_002	2579783998	2,69
07_23_007	2186152411	2,68
07_23_011	2459916886	2,70
07_23_012	2740304789	2,70

07_23_021	2484624677	2,70
07_33_004	2620228153	2,69
09_40_007	2167763267	2,70
09_40_011	2836313159	2,69
09_40_012	2222908490	2,68
09_40_019	2457575377	2,68
10_38_003	2195112794	2,68
10_38_017	2570707114	2,69
10_38_028	2596906303	2,72
10_38_029	2399591032	2,71
10_38_030	2891622038	2,72
11_39_006	2411944917	2,69
11_39_015	2273656177	2,72
12_42_001	2353047978	2,70
12_56_015	2551586997	2,70
13_44_001	2632075627	2,68
13_44_014	2727464414	2,72
13_45_023	2410272046	2,71
13_46_006_b	2442406302	2,71
13_46_011_b	2515830837	2,71
13_46_025	1581281467	2,72

Taula 10: Integració dels pic de cotinina-d3, temps de retenció dels pics i àrees de les integracions; Font: pròpia

11.06. Codi llenguatge R Total Ion Chromatogram (TIC)

```

library(xcms)
library(ggplot2)
library(magrittr)
library(plotly)
library(DT)

setwd("~/dataprocessing_urine_eclipses/Metadata")

path.mzML <- "~/dataprocessing_urine_eclipses/Metadata/QC"
mzML.files <- list.files(path.mzML, pattern= ".mzML",
                        recursive=T, full.names = T)

pheno <- phenoDataFromPaths(mzML.files)

raw_data <- readMSData(files = mzML.files,
                      pdata = new("NAnnotatedDataFrame", pheno),
                      mode = "onDisk")

# Obtinc el Total Ion Chromatogram

register(SerialParam())
TICS <- chromatogram(raw_data, aggregationFun = "sum")
newcols <- "maroon4"
names(newcols) <- levels(pData(raw_data)$class)
plot(TICS, col = newcols[raw_data$class], main = "TIC")

legend(530, 820000000, legend="QC", col= newcols, lty=1, cex=0.4, lwd = 3)

```

Figura 15. Script per extreure el Total Ion Chromatogram (TIC) dels 19 QC. Genero el vector newcols i escullo fer el TIC en lila. Li poso com a llegenda QC. Font: Llenguatge de programació R

11.07. Codi de llenguatge R Extracted Ion Chromatogram (EIC)

```
# Busco els estàndards interns (nicotina-d3, cotinina-d3 i NNAL-d3) als Quality control

# Obtinc la massa de la nicotina-d3 protonada
M <- 165.134528693
H <- 1.0078250319
M_H <- M+H

# Defineixo l'error en ppm i el rang de m/z
error <- 5 #ppm
mmu_min <- M_H-(error*M_H/1e6)
mmu_max <- M_H+(error*M_H/1e6)
mzr <- c(mmu_min, mmu_max)

raw_data |>
  filterRt(rt = c(280,293)) |>
  filterMz(mz = mzr) |>
  chromatogram(aggregationFun = "max") |>
  plot()

legend(290.1, 70000000, legend="Nicotine-d3", col= "gray", lty=1, cex=0.6, lwd = 2)

# Obtinc la massa de la cotinina-d3 protonada
M <- 179.113793248
H <- 1.0078250319
M_H <- M+H

# Defineixo l'error en ppm i el rang de m/z
error <- 5 #ppm
mmu_min <- M_H-(error*M_H/1e6)
mmu_max <- M_H+(error*M_H/1e6)
mzr <- c(mmu_min, mmu_max)
```

Figura 16. Script per extreure el Extract Ion Chromatogram (EIC) dels 19 QC. Calculo la massa monoisotòpica teòrica de la nicotina-d3, defineixo un error de massa de 5ppm i visualitzo el EIC. Calculo la massa monoisotòpica teòrica de la cotinina-d3 i defineixo un error de 5ppm de massa. Font: Llenguatge de programació R

```
raw_data |>
  filterRt(rt = c(155,170)) |>
  filterMz(mz = mzr) |>
  chromatogram(aggregationFun = "max") |>
  plot()

legend(167.7, 70000000, legend="QC", col= "gray", lty=1, cex=0.6, lwd = 3)

# Obtinc la massa del NNAL-d3 protonat
M <- 212.135256968
H <- 1.0078250319
M_H <- M+H

# Defineixo l'error en ppm i el rang de m/z
error <- 5 #ppm
mmu_min <- M_H-(error*M_H/1e6)
mmu_max <- M_H+(error*M_H/1e6)
mzr <- c(mmu_min, mmu_max)

raw_data |>
  filterRt(rt = c(100,115)) |>
  filterMz(mz = mzr) |>
  chromatogram(aggregationFun = "max") |>
  plot()

legend(112.75, 440000000, legend="QC", col= "gray", lty=1, cex=0.6, lwd = 3)
```

Figura 17. Visualitzo l'EIC de la cotinina-d3. Calculo la massa monoisotòpica teòrica del NNAL-d3, defineixo un error de massa de 5ppm i visualitzo l'EIC del NNAL-d3.

11.08. Base de dades THS

Nom	Número CAS	Fórmula molecular	Massa monoisotòpica
Nicotine	05-11-1954	C10H14N2	162.115698
Cotinine	486-56-6	C10H12N2O	176.094963
NNN	16543-55-8	C9H11N3O	177.090212
3-hydroxycotinine	34834-67-8	C10H12N2O2	192.089878
NNA	64091-90-3	C10H13N3O2	207233
NNK	64091-91-4	C10H13N3O2	207.100777
NNAL	76014-81-8	C10H15N3O2	209.116427
NAB	37620-20-5	C10H13N3O	191.23

NAT	887407-16-1	C10H11N3O	189.21
Myosmine	0532-12-07	C9H10N2	146.19
β -nicotyrine	487-19-4	C10H10N2	158.084398
2,3'-Bipyridine	581-50-0	C10H8N2	156.068748
N-Formylnormicotine	3000-81-5	C10H12N2O	176.094963
Nicotelline	0494-04-02	C15H11N3	233.095297
Methyl nicotinate	93-60-7	C7H7NO2	137.047678
N-methylnicotinamide	114-33-0	C7H8N2O	136.15
NDMA	62-75-9	C2H6N2O	74.048013
NMEA	10595-95-6	C3H8N2O	88.063663
NDEA	55-18-5	C4H10N2O	102.079313
NDPA	621-64-7	C6H14N2O	130.110613
NPyr	930-55-2	C4H8N2O	100.063663
Nmor	59-89-2	C4H8N2O2	116.058578
Npip	100-75-4	C5H10N2O	114.079313
NDBA	924-16-3	C8H18N2O	158.141913
o-toluidine	95-53-4	C7H9N	107.073499
2-aminonaphthalene	91-59-8	C10H9N	143.073499
o-Anisidine	90-04-0	C7H9NO	123.068414
Aniline	62-53-3	C6H7N	93.057849
4-aminobiphenyl	92-67-1	C12H11N	169.089149
2,6-Dimethylaniline	87-62-7	C8H11N	121.089149
9-nitroanthracene	602-60-8	C14H9NO2	223.063329
1-nitropyrene	5522-43-0	C16H9NO2	247.063329
Benz[a]anthracene	56-55-3	C18H12	228.0939
Benzo[b]fluoranthene	205-99-2	C20H12	252.0939
Benzo[j]fluoranthene	205-82-3	C20H12	252.0939
Benzo[k]fluoranthene	0207-08-09	C20H12	252.0939
Benzo[a]pyrene	50-32-8	C20H12	252.0939
Dibenz[a,h]anthracene	53-70-3	C22H14	278.10955
Dibenzo[a,i]pyrene	189-55-9	C24H14	302.10955
Dibenzo[a,e]pyrene	192-65-4	C24H14	302.10955
Indeno[1,2,3-cd]pyrene	193-39-5	C22H12	276.0939
5-Methylchrysene	3697-24-3	C19H14	242.10955
3-Ethenylpyridine	1121-55-7	C7H7N	105.057849
2-Ethenylpyridine	100-69-6	C7H7N	105.057849
2-Picoline	0109-06-08	C6H7N	93.057849
3-Picoline	108-99-6	C6H7N	93.057849
3-Ethylpyridine	536-78-7	C7H9N	107.073499
Pyridine	110-86-1	C5H5N	79.042199
Pyrrol	109-97-7	C4H5N	67.042199
Propanenitrile	107-12-0	C3H5N	55.042199
Acrylonitrile	107-13-1	C3H3N	53.026549
2-Methylfuran	534-22-5	C5H6O	82.041865
2,5-Dimethylfuran	625-86-5	C6H8O	96.057515
Benzaldehyde	100-52-7	C7H6O	106.041865

Benzene	71-43-2	C6H6	78.04695
Benzo(b)furan	271-89-6	C8H6O	118.041865
Dibenz(a,h)acridine	224-42-0	C21H13N	279.104799
Dibenzo(c,g)carbazole	194-59-2	C20H13N	267.104799
Ethylbenzene	100-41-4	C8H10	106.07825
Furan	110-00-9	C4H4O	68.026215
Cumene	98-82-8	C9H12	120.0939
m-Ethyltoluene	620-14-4	C9H12	120.0939
m,p-Xylene	179601-23-1	C9H12	120.0939
m-Xylene	108-38-3	C9H12	120.0939
p-Xylene	106-42-3	C9H12	120.0939
n-Propylbenzene	103-65-1	C9H12	120.0939
Naphthalene	91-20-3	C10H8	128.0626
o-Ethyltoluene	611-14-3	C9H12	120.0939
o-Xylene	95-47-6	C8H10	106.07825
p-Ethyltoluene	622-96-8	C9H12	120.0939
Nitrobenzene	98-95-3	C6H5NO2	123.032028
Styrene	100-42-5	C8H8	104.0626
Toluene	108-88-3	C7H8	92.0626
2,3-Butanedione	0431-03-08	C4H6O2	86.036779
2-Butanone	78-93-3	C4H8O	72.057515
Acetone	67-64-1	C3H6O	58.041865
Acetophenone	98-86-2	C8H8O	120.057515
Acrolein	0107-02-08	C3H4O	56.026215
Methacrolein	78-85-3	C4H6O	70.041865
Dichloromethane	02/09/1975	CH2Cl2	83.953356
Solanesol	13190-97-1	C45H74O	630.573967
Butane	106-97-8	C4H10	58.07825
Isobutane	75-28-5	C4H10	58.07825
n-Decane	124-18-5	C10H22	142.172151
n-Nonane	111-84-2	C9H20	128.156501
1,3-Butadiene	106-99-0	C4H6	54.04695
1,3-Pentadiene	504-60-9	C5H8	68.0626
1-Butene	106-98-9	C4H8	56.0626
1-Nonene	0124-11-08	C9H18	126.140851
1-Octene	111-66-0	C8H16	112.125201
d-Limonene	5989-27-5	C10H16	136.125201
Isoprene	78-79-5	C5H8	68.0626
Nicotine	NA	C10H14N2	162.1156985
Nicotine N'oxide	NA	C10H14N2O	178.1106131
Nicotine Isomethonium Ion	NA	C11H18N2	178.1469986
Nicotine-<U+0394> 1' (5') Iminium Ion	NA	C10H13N2	161.1078734
Cotinine Methonium Ion	NA	C11H15N2O	191.1184381
Nicotine N-glucononide	NA	C16H22N2O6	338.1477864
Cotinine N-glucononide	NA	C17H23N2O7	367.1505261
3-Pyridilacetic Acid	NA	C7H7NO2	137.0476785

Cotinine	NA	C10H12N2O	176.094963
4-(3-pyridil)-butanoic Acid	NA	C9H11NO2	165.0789786
2'-Hydroxynicotine	NA	C10H14N2O	178.1106131
Nornicotine	NA	C9H12N2	148.1000484
N'- Hydroxymethyl- norcotinine	NA	C10H12N2O2	192.0898776
Cotinine N-oxide	NA	C10H12N2O2	192.0898776
5'-Hydroxycotinine	NA	C10H12N2O2	192.0898776
Trans-3'-Hydroxycotinine	NA	C10H12N2O2	192.0898776
4-(Methylamino)-1-(3-pyridyl)-1-butanone	NA	C10H14N2O	178.1106131
4-Oxo-4-(3-pyridyl)-butanamide	NA	C9H10N2O2	178.0742276
Norcotinine	NA	C9H10N2O	162.0793129
4-Oxo-4-(3-pyridyl)-N-methylbutanamide	NA	C10H12N2O2	192.0898776
Trans-3'-Hydroxycotinine O glucuronide	NA	C16H20N2O8	368.1219656
4-(3-pyridyl)-3-butenic acid	NA	C9H9NO2	163.0633285
4-hydroxy-4-(3-pyridyl)-butanoic acid	NA	C9H11NO3	181.0738932
4-Oxo-4-(3-pyridyl)-butanoic acid	NA	C9H9NO3	179.0582431
5-(3-pyridyl)-tetrahydrofuran-2-one	NA	C9H9NO2	163.0633285
3-Vinylpyridine	NA	C7H7N	105.0578492
5'-Hydroxy-β-Nicotyrine	NA	C10H10N2O	174.0793129
β-Nicotyrine	NA	C10H10N2	158.0843983
Anabasine	NA	C10H14N2	162.1156985
Anatabine	NA	C10H12N2	160.1000484
Myosmine	NA	C9H10N2	146.0843983
β-Nornicotyrine	NA	C9H8N2	144.0687483
Metanicotine	NA	C10H14N2	162.1156985
N'-Methylanabasine	NA	C11H16N2	176.1313485
N'-Methylanatabine	NA	C11H14N2	174.1156985
N'-Methylmyosmine	NA	C10H12N2	160.1000484
2,3'-Bipyridyne	NA	C10H8N2	156.0687483
Pseudooxynicotine	NA	C10H14N2O	178.1106131
Nicotine-N-oxide	NA	C10H14N2O	178.11061
Nicotelline	NA	C15H11N3	233.0952974
N-formylnornicotine	NA	C10H12N2O	176.094963
ABT-089	NA	C11H16N2O	192.1262631
GTS-21	NA	C19H20N2O2	308.1524779
ABT-418	NA	C10H16N2O	180.1262631
A-85380	NA	C9H12N2O	164.094963
Epibatidine	NA	C11H13CIN2	208.0767261
ABT-594	NA	C9H11CIN2O	198.0559907
SIB-1508Y	NA	C12H14N2	186.1156985
RJR-2403	NA	C10H14N2	162.1156985
ABT-418 N'-oxide	NA	C9H14N2O2	182.1055277
4-(Methylnitrosamino)-1-(3-pyridyl)-1-butanol (NNAL)	NA	C16H23N3O8	385.1485147
NNAL-N-glucuronide	NA	C10H15N3O3	225.1113414

NNAL-O-glucoronide	NA	C9H13NO2	167.0946287
1-(Methylnitrosoamino)-4-(3-pyridinyl)-1,4-butanediol	NA	C10H15N3O2	209.1164267
1-(3-Pyridinyl)-1,4-butanediol	NA	C10H15N3O3	225.1113414
iso-NNAL	NA	C9H9NO2	163.0633285
NNAL-N-Oxide	NA	C10H15N3O3	225.1113414
5-(3-Pyridyl)-tetrahydro-furan-2-one	NA	C9H11NO2	165.0789786
4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK)	NA	C10H13N3O2	207.1007767
NNK-N-Oxide	NA	C10H13N3O3	223.0956913
4-[(Hydroxymethyl)nitrosoamino]-1-(3-pyridinyl)-1-butanone	NA	C10H13N3O3	223.0956913
N'-nitrosornicotine (NNN)	NA	C9H11N3O	177.090212
5-(3-Pyridyl)-2-hydroxytetrahydrofuran	NA	C9H11NO2	165.0789786
N'- Nitrosornicotine-N-b-D-glucoronide	NA	C15H19N3O7	353.1223
4-Hydroxy-1-(3-pyridyl)-1-butanone (HPB)	NA	C9H11NO2	165.0789786
4-(methylnitrosamino)-4-(3-pyridyl) butanal (NNA)	NA	C10H13N3O2	207.1007767
N'-nitrosoanabasine (NAB)	NA	C10H13N3O	191.105862
N'-nitrosoanatabine (NAT)	NA	C10H11N3O	189.090212
iso-NNAC	NA	C10H13N3O3	223.0956913
O6-Me-Gua	NA	C6H7N5O	165.0650599
(+)-Anti-bpde	NA	C20H14O3	302.0942943
Hydroquinone	NA	C6H6O2	110.0367794
1,4-Benzoquinone	NA	C6H4O2	108.0211294
Acrolein	NA	C3H4O	56.02621475
<U+03B3>-OH-PdG	NA	C14H19N5O5	337.1386187
4-ABP	NA	C12H11N	169.0891494
hexa-2,4-dienedial	NA	C6H6O2	110.0367794
Chloroacetaldehyde	NA	C2H3ClO	77.9872424
8-Hydroxyguanine	NA	C5H5N5O2	167.0443244
O4-Methylthymidine	NA	C11H16N2O5	256.1059216
8-Hydroxy-2'-deoxyguanosine	NA	C10H13N5O5	283.0916685
7-Methylguanine	NA	C6H7N5	165.0650599
DL-Nicotine-d3	NA	C10H11D3N2	NA

Taula 11. Base de dades dels composts específics del tabac (THS), aquesta llista inclou el número CAS, la fórmula molecular i la massa monoisotòpica. Font: pròpia.

11.09. Codi de llenguatge R de XCMS paràmetres optimitzats

```
## Load data
mzML.files <- list.files(path="./path2yourwd",
                        pattern= ".mzML",
                        recursive=T, full.names = T)

pheno <- phenoDataFromPaths(mzML.files)
raw_data <- readMSData(files = mzML.files,
                      pdata = new("NAnnotatedDataFrame", pheno),
                      mode = "onDisk")

## Chromatographic peak detection through centwave
cw_onlinexcms_default <- CentWaveParam(ppm = 5, peakwidth = c(5, 10),
                                       snthresh = 10, prefilter = c(3, 5000),
                                       mzCenterFun = "wMean", integrate = 1L,
                                       mzdiff = 0.01, fitgauss = FALSE,
                                       noise = 1000, verboseColumns = FALSE,
                                       roiList = list(), firstBaselineCheck = TRUE,
                                       roiScales = numeric())

xdata <- findChromPeaks(raw_data, param = cw_onlinexcms_default)

## Alignment
xdata <- adjustRtime(xdata, param = ObiwrapParam(binSize = 1))
```

Figura 18. Escullo el meu directori de treball i llegeixo les dades corresponents als tres grups (exposats, no exposats i QC). Estableixo els paràmetres per detectar pics cromatogràfics, els detecto i els alineo. Font: Llenguatge de programació R.

```
## Correspondence
pdp_onlinexcms_default <- PeakDensityParam(sampleGroups = pheno$class,
                                           minFraction = 0.8,
                                           bw = 5, binSize = 0.005)

xdata <- groupChromPeaks(xdata, param = pdp_onlinexcms_default)

## Fill peaks
xdata <- fillChromPeaks(xdata)

## Extract feature definitions
feature.info <- as.data.frame(featureDefinitions(xdata))

## Extract feature matrix
fmat <- featureValues(xdata, value = "maxo")
```

Figura 19. Agrupo els pics entre si per formar les *features*. Busco els valors dels pics que s'han perdut durant l'anàlisi. Extreco la matriu de *features* (feature.info) i la matriu d'intensitat (fmat). Font: Llenguatge de programació R.

11.10. Codi de llenguatge R filtrat de les dades de metabolòmica

```
load("~/Camilla/R_files/Metadada/xdata_parametres_modificats.RData")
feature.info <- as.data.frame(featureDefinitions(xdata))
setwd("~/Camilla/R_files/Metadada")
library(ggplot2)
library(plotly)

##Genero la matriu sobre la que treballaré i indico les columnes i els grups
D <- as.data.frame(featureValues(xdata, value = "maxo", method = "maxint"))
colnames(D) <- row.names(pData(xdata))
c1 <- as.character(pData(xdata)$class)

#####PRIMER NORMALITZO LA MATRIU#####
library(readr)
Urianalisi <- read_delim("Urianalisi.csv",
                        delim = ";", escape_double = FALSE, trim_ws = TRUE)

D.norm.t <- apply(D, 1, function(x) x/Urianalisi$SPECIFIC_GRAVITY)
D.norm <- t(D.norm.t)
```

Figura 20. Escullo el meu directori de treball, genero la matriu D i la normalitzo amb la meua densitat específica. Font: Llenguatge de programació R.

```
#####FAIG UN PCA PRELIMINAR #####
D.normPCA <- as.data.frame(t(apply(D.norm.t, 1, function(x) (x/max(x, na.rm = TRUE))))))
D.normPCA[is.na(D.normPCA)] <- 0

pca_data <- prcomp(D.normPCA,retx=TRUE, center=TRUE, scale=FALSE)
summary(pca_data)$importance[,c(1:3)]

scores <- data.frame(pca_data$x[, c("PC1", "PC2")])
scores$class <- classe
scores$lab <- rownames(D.normPCA)
scores.plot <- ggplot(data = scores,
                      aes(x = PC1,
                           y = PC2, colour = classe,
                           label=lab)) +
  geom_point(alpha = I(0.7), size = 0.5) +
  geom_hline(yintercept = 0)+
  geom_vline(xintercept = 0)+
  xlab(paste("PC1 (", round(summary(pca_data)$importance[2,1], 2) * 100, "%)"))+
  ylab(paste("PC2 (", round(summary(pca_data)$importance[2,2], 2) * 100, "%)"))+
  stat_ellipse() +
  theme_bw()

ggplotly(scores.plot)

### Observo que tinc outliers i vaig a la raw_data per veure si és un problema analític, les mostres
### 03_17_055, 03_17_003 i 04_18_003 no tenen internal standard, això implica que no sel's hi
### va afegir la solució extractant, les elimino
```

Figura 21. Faig un PCA exploratori i observo a les dades crues que tres de les meves mostres no tenen estàndards interns i per tant no tenen solució extractant. Font: Llenguatge de programació R.

```
which (colnames (D.norm)== "03_17_055")
which (colnames (D.norm)== "03_17_003")
which (colnames (D.norm)== "04_18_012")

D.norm <- D.norm [,-c(109,113,116)]
D.norm.t <- t(D.norm)
classe.outliers <- classe [-c(109,113,116)]

### Aquest pas em calcula la intensitat mediana de cada grup
###He d'eliminar també les columnes de la classe de les mostres que hem eliminat abans

median.intensities <- as.data.frame(t(apply(D.norm,
1, function(x) tapply(x, classe.outliers, median, na.rm = TRUE))))

##Elimino el QC per calcular la mitjana de les intensitats
median.intensities$QC <- NULL

### Estableixo un valor llindar mínim d'intensitat
thresholdvalue <- 40000

### Ara el que faig es dir-li que si en almenys un grup tinc d'una feature un valor de >10000
### me la quedo, si a un grup tinc 9000 pero a l'altre tinc 11000, me la quedo, ja que el que
### vull és veure diferències entre els grups, llavors segurament d'alguna feature que sigui
### diferent, en un grup tindre menys del llindar i en l'altre més

idx_i <- names(which(apply(median.intensities,
1, function(x) any(x>thresholdvalue)==TRUE)==TRUE))

cat(paste(length(idx_i), "out of", dim(median.intensities)[1]),
"above the threshold intensity value")
```

Figura 22. Busco les mostres a la columna de la meua matriu i les elimino de l'estudi. Començo amb el filtrat per intensitat, calculo la intensitat mediana dels grups, elimino els QC per aquest càlcul i estableixo com a llindar d'intensitat 40000. Amb aquest filtrat 11662 de les 16206 features inicials passen en almenys un dels grups el llindar d'intensitat Font: Llenguatge de programació R.

```

### Ara em donarà com output quantes de les meves features en almenys un dels grups
### supera el llindar de 10000 d'intensitat

# Anàlisi exploratòria de dades per avaluar la variabilitat analítica

## Primer càlculo el RSD%
RSD <- function(x){100*sd(x, na.rm = TRUE)/mean(x, na.rm = TRUE)}

## Genero un vector amb els QC (QChits) i un altre amb les mostres (Samples_hits)
QChits <- which(classe.outliers == "QC")
Samples_hits <- which(classe.outliers=="Exposats_SHS_THS" | classe.outliers=="No_exposats_SHS_THS")

## Càlculo l'RSD de les mostres (RSD_Samples) i el dels QC (RSD_QC)-----
RSD_samples <- apply(D.norm[,Samples_hits],1, RSD)
RSD_QC <- apply(D.norm[,1, RSD)

## Elimino aquelles features que tinguin una major variació analítica (QC) que biològica
idx_qc <- names(RSD_samples)[which(RSD_QC< RSD_samples)]

cat(paste(length(idx_qc), "out of", dim(median.intensities)[1]),
    "hold higher biological than analytical variation")

D.norm_filtrada <- D.norm[intersect (idx_1,idx_qc), -c(grep("QC", classe.outliers))]
D.norm_filtrada_t <- t(D.norm_filtrada)

#####FAIG UN PCA PER VEURE SI SEGUEIXO TENINT ELS OUTLIERS#####

classe_outlier_senseQC <- classe.outliers[-c(grep("QC", classe.outliers))]
D.normPCA <- as.data.frame(t(apply(D.norm_filtrada_t, 1, function(x) (x/max(x, na.rm = TRUE))))))
D.normPCA[is.na(D.normPCA)] <- 0

```

Figura 23. Començo amb el filtrament per variabilitat biològica > variabilitat analítica. Calculo el %RSD dels 19 QC i de les 143 mostres. Amb aquest filtrament 6667 features tenen més variabilitat biològica que analítica, la resta de features s'eliminen. Començo fent el PCA de la matriu filtrada. Font: Llenguatge de programació R.

```

pca_data <- prcomp(D.normPCA,retx=TRUE, center=TRUE, scale=FALSE)
summary(pca_data)$importance[,c(1:3)]

scores <- data.frame(pca_data$x[, c("PC1", "PC2")])
scores$class <- classe_outlier_senseQC
scores$lab <- rownames(D.normPCA)
scores.plot <- ggplot(data = scores,
  aes(x = PC1,
      y = PC2, colour = class,
      label=lab)) +
  geom_point(alpha = I(0.7), size = 0.5) +
  geom_hline(yintercept = 0)+
  geom_vline(xintercept = 0)+
  xlab(paste("PC1 (%)", round(summary(pca_data)$importance[2,1], 2) * 100, "%"))+
  ylab(paste("PC2 (%)", round(summary(pca_data)$importance[2,2], 2) * 100, "%"))+
  stat_ellipse() +
  theme_bw()

ggplotly(scores.plot)

## Comprovant la metadata que tinc i els outliers del PCA, decideixo eliminar les mostres
##02_11_429, 02_29_238 i 02_09_439 (menjar xocolata el dia anterior),02_28_399, 01_04_182,
##02_12_368, 01_05_155, 01_27_071 (presenten leucòcits alterats), la mostra 02_32_157
##(presenta hemoglobina alterada) i la mostra 03_17_045 (menjar peix el dia anterior)

## En primer lloc, he de buscar a quines files estan aquestes mostres
which (colnames (D.norm_filtrada)== "02_11_429")
which (colnames (D.norm_filtrada)== "02_29_238")
which (colnames (D.norm_filtrada)== "02_28_399")
which (colnames (D.norm_filtrada)== "01_04_182")
which (colnames (D.norm_filtrada)== "02_12_368")

```

Figura 24. Faig el PCA i observo que tenim outliers, de manera que reviso la metadata del urianàlisi i els aliments que van ingerir el dia anterior que alteren el metaboloma (xocolata, peix, bacu) i elimino les mostres que van ingerir aquests aliments o presenten leucòcits o hemoglobina en sang. Font: Llenguatge de programació R.

```

which (colnames (D.norm_filtrada)== "02_12_368")
which (colnames (D.norm_filtrada)== "01_05_155")
which (colnames (D.norm_filtrada)== "02_32_157")
which (colnames (D.norm_filtrada)== "01_27_071")
which (colnames (D.norm_filtrada)== "03_17_045")
which (colnames (D.norm_filtrada)== "02_09_439")

## Elimino els outliers que tinc a les columnes
D.norm_filtrada_outliers <- D.norm_filtrada[,-c(23,28,59,62,73,81,87,91,97,110)]
D.norm_filtrada_outliers[is.na(D.norm_filtrada_outliers)] <- 0

D.norm_filtrada_outliers_t <- t(D.norm_filtrada_outliers)
cl_sense_QC_outliers <- classe_outlier_senseQC [-c(23,28,59,62,73,81,87,91,97,110)]

## Filtro les dades per aquelles que també estan a l'HMDB, també li poso la informació
## sobre les masses adductes més possibles per a que també ho tingui en compte

## Llegir la informació sobre els adductes positius (hem treballat en mode positiu)
positive <- read.table("./Positive_ESI.txt", header = T, stringsAsFactors = F)

## Read HMDB database
hmdb <- read.table("./HMDB_17_09.txt", header = T, sep = "\t", stringsAsFactors = F,
  fill = T, quote = "")

## Genero un data frame de les meves intensitats i li aplico els filtres d'intensitat i de variància analítica
RES <- cbind.data.frame(median.intensities)
RES.sig <- RES[intersect (idx_1,idx_qc),]
input.mz <- feature.info[rownames(RES.sig), "mzmed"]
input.rt <- feature.info[rownames(RES.sig), "rtmed"]

```

Figura 26. Busco les features filtrades a una base de dades de la HMDB per eliminar les features que no s'hagin trobat anteriorment en humans (incloent les features provinents de la columna). Font: Llenguatge de programació R.

```

# Estableixo un error de 5 ppm perquè utilitzo un Orbitrap
ppm <- 5

hmdbhits <- lapply(input.mz, function(x) {
  search_vect <- (x + positive$AdductMass) # here we are using positive ionisation
  h <- lapply(search_vect, function(x) {
    mass_range <- c(x - ppm * (x/1e+06), x + ppm * (x/1e+06))
    a <- which(hmdb$MonoisotopicMass > mass_range[1] & hmdb$MonoisotopicMass <
      mass_range[2])
  })

  h2 <- sapply(1:length(h), function(x) {
    y <- h[[x]]
    if (length(y) > 0) {
      r <- paste(hmdb$name[y], sep = "", collapse = ";")
    } else {
      r <- "NA"
    }
    return(r)
  })

  return(h2)
})

hmbdhits <- do.call("rbind", hmbdhits)
colnames(hmbdhits) <- positive$Adduct

RESULTS2MSMS <- cbind.data.frame(RES.sig, input.mz, input.rt, hmbdhits)

pos <- NULL
for(i in 1:nrow(hmbdhits)) {
  if (any(hmbdhits[i,] != "No hit")) {
    pos <- c(pos, i)
  }
}

D.norm_filtrada_outliers_hmdb <- D.norm_filtrada_outliers[pos,]

write.csv(D.norm_filtrada_outliers_hmdb, file="D.norm_filtrada_outliers_hmdb.csv")
save(D.norm_filtrada_outliers_hmdb, file="D.norm_filtrada_outliers_hmdb.RData")

```

Figura 26. Per últim, comparo les meves *features* amb les presents a la HMDB i elimino de la meua matriu aquelles *features* no presents, de manera que passen el filtre 2032 *features*. Font: Llenguatge de programació R.

11.11. Codi de llenguatge R del boxplot nicotina i hidroxicotinina

```

library(xcms)
library(ggplot2)
library(magrittr)
library(plotly)
library(DT)

setwd("~/dataprocessing_urine_eclipses/Metadadata")

path.mzML <- "~/dataprocessing_urine_eclipses/Metadadata/QC"
mzML.files <- list.files(path.mzML, pattern = ".mzML",
  recursive=T, full.names = T)

pheno <- phenoDataFromPaths(mzML.files)

raw_data <- readMSData(files = mzML.files,
  pdata = new("NAnnotatedDataFrame", pheno),
  mode = "onDisk")

load("~/Camilla/R_files/Metadadata/xdata_parametres_modificats.RData")
D <- as.data.frame(featureValues(xdata, value = "maxo", method = "maxint"))

# Faig un boxplot de la nicotina
boxplot(as.numeric(D["FT02934",])~pData(xdata)$class, xlab="", ylab="",
  main="Nicotine")

# Faig un boxplot de l'hidroxicotinina
boxplot(as.numeric(D["FT04395",])~pData(xdata)$class, xlab="", ylab="",
  main="OH-cotinine")

```

Figura 27. Escullo el meu directori de treball i lleigeixo els arxius en format .mzML. Carrego el objecte obtès a xcms (xdata) i genero la matriu de *features* (D). Faig un boxplot de la nicotina (FT02934) i de l'hidroxicotinina (FT04395). Font: Llenguatge de programació R.

ANNEX 2

FITXA DE SEGUIMENT DEL TUTOR/A del TFG

Nom i Cognoms de l'Alumne/a: CÈLIA TORRES VILANOVA

Nom i Cognoms del Tutor/a: JAVIER CAPILLA LUQUE

Data de la entrevista amb l'alumne: 03/03/2022


Recomanacions durant el seguiment: Entrevista virtual. Recomanacions en quant a l'estructuració del treball i contingut dels apartats/subapartats. S'ha discutit la possibilitat d'incloure la detecció de components immunes a l'orina dels pacients inclosos a l'estudi. Es buscarà bibliografia per estudiar la viabilitat i interès. S'han establert recomanacions en quant a com estructurar el treball escrit i estil de la redacció.

Observacions: L'alumna té una bona disposició i capacitat per fer propostes ben raonades

Observacions Darrera revisió (25/05/2022):

El treball està ben estructurat. A falta de completar l'apartat de resultats, la redacció és molt adequada, precisa i ben estructurada. S'anima a l'alumna a continuar aquesta dinàmica amb els apartats restants. Es recomana evitar la inclusió de protocols a la secció metodològica i incloure'ls com annexes.

Signatura del Tutor/a



Signatura del Alumne/a



Tarragona a 26 de maig de 2020