

**Pol Romera Viñuelas**

**Composició multi-òmica i correlacions de l'alè i la saliva**

**Treball Fi de Grau  
dirigit per la Dra. Raquel Cumeras**

**Grau en Enginyeria Biomèdica**



UNIVERSITAT ROVIRA I VIRGILI

**Tarragona  
2023**

# Índex

Resum .....	3
Agraïments.....	4
1 Introducció.....	5
2 Objectius .....	8
3 Metodologia .....	9
3.1 Fonts de composició completa de la saliva i l'alè.....	9
3.2 Metabòlits.....	10
3.3 Proteïnes i gens.....	11
3.4 Elements químics.....	12
4 Noves bases de dades de la composició de la saliva i l'alè .....	13
4.1 Creació de la base de dades .....	13
4.2 Atributs de la base de dades .....	13
4.3 Recull d'atributs.....	15
4.4 Desenvolupament del codi.....	16
5 Base de dades Life Fluids de la Saliva .....	18
5.1 Life Fluids SALIVA metabòlits.....	18
5.2 Life Fluids SALIVA proteïnes .....	20
5.3 Life Fluids SALIVA elements químics.....	22
6 Base de dades Life Fluids de l'alè .....	23
6.1 Life Fluids ALÈ metabòlits.....	23
6.2 Life Fluids ALÈ proteïnes .....	25
6.3 Life Fluids ALÈ elements químics.....	26
7 Correlacions entre les dades Life Fluids de saliva i d'alè.....	27
8 Discussió.....	28
8.1 Creació, atributs i curació .....	28
8.2 Metabòlits.....	29
8.3 Proteïnes .....	31
8.4 Elements químics.....	32
8.5 Correlacions Saliva-Alè.....	33
8.6 Limitacions.....	34
8.7 Disponibilitat i protecció de les dades generades .....	34
9 Conclusions .....	35
10 Annex .....	36
11 Referències .....	37

## Resum

L'objectiu principal d'aquest treball de fi de grau és desenvolupar una base de dades unificada que incorpori compostos presents a dos biofluids humans: la saliva i l'alè. Aquest estudi pretén incrementar la quantitat de metabòlits, proteïnes i gens associats, així com elements químics identificats fins ara. Els objectius específics inclouen la identificació de fonts disponibles per a la composició de saliva i alè, l'organització de la nova base de dades i la correlació dels components comuns entre tots dos biofluids. La composició de saliva i alè es va classificar en tres categories: metabòlits, proteïnes i gens associats, i elements químics. Per construir la base de dades, es va utilitzar informació de bases de dades existents, literatura i repositoris. Per superar les discrepàncies a la nomenclatura, es van utilitzar identificadors específics: InChIKey per a metabòlits i elements químics, i Uniprot per a proteïnes. També s'hi van afegir identificadors públics àmpliament utilitzats, com ara PubChem, HMDB, LipidMaps, KEGG i Uniprot. La nova base de dades "Life Fluids" inclou 6727 entrades entre els compostos de la saliva; i 1773 entre els de l'alè. Aquest treball representa un pas significatiu cap a la consolidació de la informació existent sobre els components de la saliva i l'alè humà. La nova base de dades unificada proporciona un recurs valuós per als investigadors que estudien aquestes àrees, permetent una anàlisi més profunda i comparativa d'aquests biofluids. A més, la metodologia emprada per superar les discrepàncies de nomenclatura es pot aplicar en altres estudis que requereixin la unificació d'informació procedent de diferents fonts.

## **Agraïments**

Volia començar agraint a la Dra. Raquel Cumeras, PhD i a la Maria Llambrich, estudiant doctorat PIPF, per confiar amb mi per poder fer el meu TFG amb el seu equip, donar-me les pautes i les eines per dur a terme aquest treball.

M'agradaria tenir un menció especial a Hector Villalba, tècnic de bases de dades, per supervisar el treball, ajudar-me en els dubtes i guiar-me.

Sol tinc paraules de gratitud a tot l'equip a tot l'equip de MILAB i els hi desitjo el major dels èxits en tots els presents i futurs projectes.

Per acabar, no m'agradaria oblidar donar gracies a la meva família, companys i amics pel suport i no deixar que em rendeixi durant aquest període que ha durat l'època universitària, i per estar al meu costat des de sempre.

## 1 Introducció

La saliva, un fluid biològic complex i multifacètic, acull una gamma impressionant de biomolècules. Dins la seva composició trobem ADN, ARN, proteïnes, metabòlits i una rica microbiota. Aquest còctel molecular brinda una immensitat de pistes diagnòstiques, que poden ser explotades per al cribratge i diagnòstic de múltiples malalties. A més, la recol·lecció de saliva ofereix avantatges significatius: és un procediment senzill, no invasiu i el seu emmagatzematge és convenientment pràctic. Aquestes característiques situen la saliva com un candidat ideal per a la realització de biòpsies líquides, implantant-la com un fluid biològic idoni per al diagnòstic[1].

D'altra banda, l'alè humà també és una font d'informació mèdica. Conté compostos orgànics volàtils (VOCs), els quals són produïts de forma endògena com a resultat de diversos processos fisiològics. Aquests compostos també es poden originar de l'activitat dels bacteris presents al tracte gastrointestinal i a les vies respiratòries, o ser el resultat del contacte amb l'entorn[2]. Com la saliva, l'alè també s'utilitza per estudiar malalties. Un exemple és el test d'alè per a la detecció de l'*Helicobacter Pylori*[3], un bacteri que s'ha associat a diverses malalties gàstriques. A més, s'han realitzat diversos estudis que utilitzen l'anàlisi de l'alè per detectar el càncer de pulmó[4] i per al monitoratge no invasiu de la diabetis[5].

Actualment, per avaluar l'estat de salut d'una persona, utilitzem mètodes més avançats coneguts com a tecnologies "òmiques"[6] (veure Figura 1). Les mesures "òmiques" involucren la identificació i quantificació de biomolècules amb l'objectiu global d'inferir l'estat fisiològic d'un organisme en base al tipus molecular, la localització i qualsevol canvi en l'abundància. Les "òmiques" basades en molècules es poden subdividir en genòmica, transcriptòmica, proteòmica i metabolòmica, totes elles proporcionant informació valuosa per comprendre l'estat biòtic complet dels sistemes biològics.

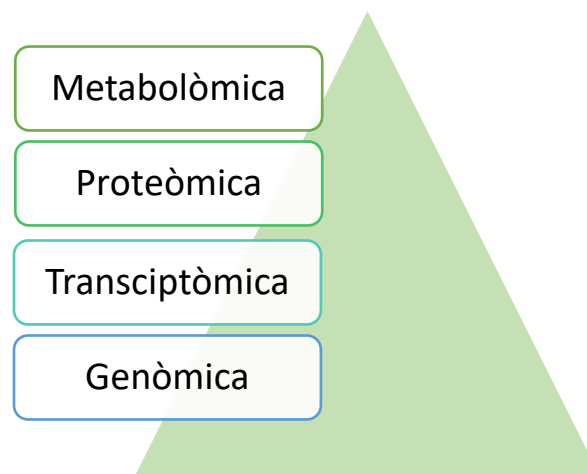


Figura 1. Representació esquemàtica de les ciències òmiques i la seva interconnexió a la biologia de sistemes. Començant des de la base amb la genòmica, seguit per la transcriptòmica i la proteòmica, fins a arribar a la metabolòmica al cim. Cada nivell representa un nivell addicional de complexitat i detall a la nostra comprensió dels processos biològics.

La genòmica, s'enfoca a l'anàlisi integral de l'estructura, la funció i l'expressió dels gens que conformen el genoma d'un organisme [7]. Aquest camp de recerca té com a objectiu comprendre la informació genètica continguda a l'ADN i com es tradueix en les característiques fenotípiques i els processos biològics d'un organisme [8]. La genòmica es basa en una varietat

de tècniques i tecnologies innovadores que permeten seqüenciar i analitzar l'ADN, identificar gens individuals i determinar-ne l'organització al genoma [9]. A més, s'estudia l'activitat dels gens a nivell molecular i cel·lular, investigant com s'expressen i es regulen [10].

La proteòmica és l'estudi integral de les proteïnes en un sistema biològic, proporcionant informació valuosa sobre la seva estructura, funció i expressió. S'utilitzen tècniques com l'espectrometria de masses per identificar i quantificar proteïnes i la bioinformàtica per a l'anàlisi de dades [11], [12].

La metabolòmica és un camp en evolució ràpida de la ciències "òmiques" que se centra en l'anàlisi integral de metabòlits i processos metabòlics en sistemes biològics. El desenvolupament continu de les tecnologies de la metabolòmica ha donat lloc a una gamma creixent d'aplicacions biomèdiques. En particular, la metabolòmica té un paper crucial en el diagnòstic de malalties, proporcionant informació sobre els mecanismes de la malaltia, descobrint nous objectius de fàrmacs, personalitzant els tractaments farmacològics i supervisant l'eficàcia del tractament[13]. Com a resultat, la metabolòmica és una gran promesa per avançar en la medicina de precisió i millorar els resultats dels pacients.

A mesura que el coneixement s'aprofundeix, la diferenciació entre les "òmiques" es fa més complexa. Com a resultat, emergeix el concepte de biologia de sistemes. Aquest enfocament en la recerca biomèdica busca comprendre la imatge global d'un organisme, teixit o cèl·lula, integrant les seves diverses parts [14]. Aquesta aproximació contrasta notablement amb dècades de biologia reduccionista, que descomponia les parts en elements individuals. La identificació sistemàtica dels elements (com ara l'ADN, ARN, proteïnes, metabòlits, etc.) que formen part d'un sistema biològic constitueix la base essencial per al desenvolupament de la biologia de sistemes.

En les ciències "òmiques" tenen aplicacions àmplies en diversos fluids i teixits corporals. Els biofluids humans es poden classificar en classes diferents segons la seva funció o localització principal, tal com s'indica a la Taula 1, on apareixen els principals biofluids. Es caracteritzen per la seva capacitat d'adaptació a la pressió externa i no tenen una forma fixa. Tot i que la majoria dels biofluids humans estan en forma líquida, alguns biofluids, com l'alè, existeixen en estat gasós.

Taula 1. Principals biofluids humans i les seves funcions

Biofluid	Funció principal
Sang	Transport de nutrients, oxigen i residus
Orina	Eliminació de productes de desfecció i toxines
Saliva	Digestió inicial dels aliments i lubricació
Alè	Reflexió de metabòlits volàtils i estat metabòlic de l'organisme
Suor	Regulació de la temperatura corporal
Femta	Eliminació de productes de desfecció del tracte digestiu
Llàgrimes	Humectació i protecció dels ulls

També podem trobar biofluids com ara la limfa que juga un paper crucial en el sistema limfàtic, transportant nutrients, hormones i residus metabòlics arreu del cos. Els líquids gastrointestinals, com el suc gàstric, la bilis, el suc pancreàtic i les flatulències; els biofluids nasofaríngics, com l'esput, el moc; els biofluids reproductius, com el flux vaginal i el semen; els biofluids de la gestació, com el líquid amniòtic, el líquid contingut a les cavitats embrionàries transitòries i la llet materna; els biofluids de l'oïda, com el cerumen, l'endolimfa i la perilimfa; el líquid cefalorraquídi; les bombolles; el pus; els líquids de la cavitat, com el pleural, el peritoneal i el pericàrdic; i els biofluids articulars, com el líquid sinovial i la bossa sinovial[15].

S'han desenvolupat diverses bases de dades (BBDD) per gestionar les quantitats massives de dades generades per les ciències "òmiques". Algunes de les més rellevants inclouen la Base de Dades del Metaboloma Humà (HMDB) [16], Exocarta [17], l'Enciclopèdia de Kyoto de Genes i Genomes [18] i moltes altres. Tot i això, l'acumulació constant de nova informació ha generat un entorn desorganitzat i ha obligat els investigadors a mantenir-se al dia tant amb diverses bases de dades com amb literatura científica i fins i tot repositoris, que sovint contenen informació duplicada. A més, la manca de classificació per fluids biològics en la majoria d'aquestes bases de dades dificulta l'accés i l'ús eficient de la informació. Altrament, cada any es publiquen nombrosos articles científics sobre nous enfocaments en les ciències "òmiques", la qual cosa dificulta la identificació de les investigacions rellevants. Amb aquesta finalitat, el projecte Life Fluids© ( [www.lifefluids.com](http://www.lifefluids.com) ) té com a objectiu crear una plataforma específica per a biofluids que abordi aquestes limitacions. El projecte se centra en la creació i millora de bases de dades de composició per a cada biofluid humà, on tots els components (gens, proteïnes, metalls, metabòlits, etc.) estiguin disponibles de forma lliure i que es pugui descarregar.

## 2 Objectius

L'objectiu principal d'aquest treball de fi de grau és crear una base de dades nova, de dos biofluids humans com són la saliva i l'alè, que unifiqui tots els compostos registrats en bases de dades anteriors. S'espera que aquest treball condueixi a un augment en el nombre de metabòlits, proteïnes i els gens associats, així com d'elements químics registrats fins ara.

Els objectius específics són els següents:

**Objectiu 1:** Identificar les fonts disponibles de composició de la saliva i l'alè, i els atributs principals comunament utilitzats a cada camp científic per classificar metabòlits, proteïnes i elements químics.

**Objectiu 2:** Centralitzar, organitzar i guarir la nova base de dades de la saliva i l'alè.

**Objectiu 3:** Correlacionar els components en comú de la saliva i l'alè.



### 3 Metodologia

#### 3.1 Fonts de composició completa de la saliva i l'alè

La composició de la saliva i l'alè es pot dividir en tres categories principals: metabòlits, proteïnes i els gens associats, i elements químics. Per crear les bases de dades detallades de composició dels dos biofluids, es va utilitzar informació proporcionada per bases de dades existents, literatura i repositoris, tal com es detalla a la Taula 2 i Taula 3. Els identificadors dels estudis en els repositoris, es detallen més endavant, en les Taula 5 i Taula 6.

Taula 2. Fonts de composició de la saliva

Categoria	Base de dades	Literatura	Repositoris
<b>Metabòlits</b>	<b>HMDB:</b> Human Metabolome DB [16] <b>BV:</b> BinVestigate[19] <b>MDB:</b> MarkerDB[20]	<b>LTMB0001:</b> Drabinska et al., 2021[21] <b>LTMB0068:</b> Pereira et al., 2020[22] <b>LTMB0069:</b> Gardner et al., 2020[23] <b>LTMB0070:</b> Huang et al., 2023[24]	<b>LTMB0071-0076:</b> Metabolomics Workbench[25] <b>LTMB0077-0081:</b> MetaboLights[26]
<b>Proteïnes</b>	<b>EXOC:</b> ExoCarta[17] <b>VSCP:</b> Vesiclepedia[27] <b>HBFP:</b> Human Body Fluid Proteome[28] <b>HSP:</b> Human Salivary Proteome[29]	<b>LTPR0001:</b> Lacombe et al., 2018[30]	
<b>Elements químics</b>	<b>HMDB:</b> Human Metabolome DB[16] <b>MDB:</b> MarkerDB[20]	<b>LTMB0001:</b> Drabinska et al., 2021[21]	<b>LTMB0075-0076:</b> Metabolomics Workbench[25]

Taula 3. Fonts de composició de l'alè

Categoria	Base de dades	Literatura	Repositoris
<b>Metabòlits</b>	<b>HMDB:</b> Human Metabolome DB[16] <b>MDB:</b> MarkerDB[20] <b>KPDB:</b> Knapsack[31]	<b>LTMB0001:</b> Drabinska et al., 2021[21] <b>LTMB0082:</b> Sharma et al., 2023[32]	<b>LTMB0083-0084:</b> Metabolomics Workbench[25] <b>LTMB0085:</b> MetaboLights[26]
<b>Proteïnes</b>	<b>HBFP:</b> Human Body Fluid Proteome[28]	<b>LTPR0001:</b> Lacombe et al., 2018[30]	
<b>Elements químics</b>		<b>LTMB0001:</b> Drabinska et al., 2021[21]	<b>LTMB0084:</b> Metabolomics Workbench[25]

Per a cada tipus de compost, es van fer servir diferents estratègies per buscar el contingut de la saliva i l'alè.

### 3.2 Metabòlits

Per als metabòlits, en primer lloc, incloem els compostos disponibles a les bases de dades públiques, en segon lloc, incloem revisions de literatura i articles rellevants que proporcionessin llistes detallades de compostos, i finalment incloem aquells trobats en repositoris de dades públiques.

**Bases de dades de metabòlits:** S'han realitzat cerques de metabòlits en les 4 bases de dades online més importants:

La "Human Metabolome Database" o HMDB ( <https://hmdb.ca> ) ha proporcionat informació de referència exhaustiva sobre els metabòlits humans i les seves propietats biològiques, fisiològiques i químiques associades des de 2007. En els últims 15 anys, la HMDB ha crescut i evolucionat significativament per satisfer les necessitats de la comunitat de la metabòmica i respondre als continus canvis en la tecnologia d'internet i la computació[33]. Es va llançar amb 2.180 metabòlits[34], i amb la darrera actualització el 2021 ha augmentat fins a 217.900[33]. A més, la HMDB proporciona l'opció de filtratge per fluid. Per això treballarem a partir d'aquesta base de dades com a punt de partida per als components de la saliva i l'alè.

El "BinVestigate" o BV ( <https://binvestigate.fiehnlab.ucdavis.edu/> ), és una aplicació web utilitzada per explorar els compostos trobats a BinBase[35], la qual conté informació solament de la saliva. BinVestigate acull milers de mostres úniques provinents de nombrosos experiments, totes analitzades mitjançant cromatografia de gasos acoblada a espectrometria de masses. Cada mostra està caracteritzada per propietats com ara l'espectre de masses, l'índex de retenció (RI), la massa de quantificació, una llista de masses úniques i un número d'identificació únic [36]. A diferència de HMDB, BinVestigate proporciona dades de múltiples espècies, no només de *Homo sapiens*. Pel que fa la saliva, sol es va trobar dues combinacions de metadades possibles, especificades a la Taula 4.

Taula 4. Possibles combinacions de la BinVestigate.

Espècie	Òrgan
Human	Saliva
Human	saliva

La "MarkerDB" o MDB ( <https://markerdb.ca/> ) és una base de dades en línia que recopila informació sobre metabòlits de la saliva i l'alè, incloent-hi tant biomarcadors clínics coneguts com un conjunt selecte de biomarcadors preclínic. El seu objectiu és proporcionar un recurs centralitzat que concentri aquestes dades en un sol lloc[37]. La base de dades conté 142 biomarcadors de proteïnes, 1,089 biomarcadors químics, 154 biomarcadors de cariotip i 26,374 marcadors genètics[20].

La "KNAPSAcK DB" o KPDB ( [http://www.knapsackfamily.com/knapsack\\_core/top.php](http://www.knapsackfamily.com/knapsack_core/top.php) ), en aquesta base de dades trobem informació de metabolits de l'alè. KNAPSAcK Core DB és una base de dades àmpliament utilitzada en el camp de la química i la biologia de productes naturals. Aquesta base de dades recopila i organitza informació sobre compostos químics provinents de fonts naturals com a plantes, fongs i microorganismes. La base de dades ofereix eines i funcions avançades que faciliten l'anàlisi i l'exploració de les dades, cosa que contribueix al descobriment i desenvolupament de compostos bioactius[38].

**Literatura de metabòlits:** Per recopilar metabòlits dels dos biofluids estudiats a partir d'estudis de metabòmica, hem utilitzat motors de cerca públics com SCOPUS, PubMed i Web of Science. S'han seleccionat diversos articles amb metadades públiques disponibles. Aquells seleccionats són principalment bases de dades que no compten amb un lloc web específic. Per la saliva s'ha seleccionat quatre articles: Drabinska et al., 2021[21], Pereira et al., 2020[22],

Gardner et al., 2020[23], Huang et al., 2023[24]. Per l'alè se n'han seleccionat dos: Drabinska et al., 2021[21] i Sharma et al., 2023[32].

**Repositoris de metabòlits:** S'han incorporat resultats de metabòlits d'estudis disponibles en dos repositoris en obert.

El "Metabolomics Workbench" ( <https://www.metabolomicsworkbench.org/> ), és un repositori internacional de dades i metadades de metabolòmica. És possible fer filtrats per espècie i per fonts de mostra. En el nostre cas, seleccionem l'espècie "Human" i les fonts de mostra "Saliva" i "Breath". Aquesta cerca va donar 6 estudis en què la saliva va ser utilitzada i un total de 2 estudis en l'alè. Els noms dels estudis incorporats es troben a la Taula 5.

Taula 5. Estudis del repositori Metabolomics Workbench amb les combinacions Humà-Saliva i Humà-Alè (Breath en anglès).

Espècie	Fonts de mostra	Estudi ID	Referencia
Human	Saliva	ST000418	LTMB0071 [39]
		ST000496	LTMB0072 [40]
		ST000588	LTMB0073 [41]
		ST001905	LTMB0074 [42]
		ST002265	LTMB0075 [43]
		ST002328	LTMB0076 [44]
	Breath	ST000587	LTMB0083 [45]
		ST000883	LTMB0084 [46]

El "MetaboLights" ( <https://www.ebi.ac.uk/metabolights/> ), és el primer repositori de bases de dades d'accés obert dissenyat per a la investigació en metabolòmica a l'Institut Europeu de Bioinformàtica (EMBL-EBI). Permet als investigadors compartir dades i metadades d'estudis de metabolòmica a diferents plataformes i espècies[47]. Pel que fa al número d'estudis trobats on s'ha utilitzat els biofluids estudiats, ha sigut: 6 estudis de la saliva i 1 sol estudi sobre l'alè. Els noms dels fitxers corresponents es troben a la Taula 6.

Taula 6. Estudis del Metabolomics Workbench amb les combinacions Homo Sapiens-Saliva i Homo Sapiens-Alè (Breath en anglès).

Organisme	Part de l'organisme	Estudi ID	Referencia
Homo Sapiens	Saliva	MTBLS2967	LTMB0077 [48]
		MTBLS2108	LTMB0078 [49]
		MTBLS760	LTMB0056 [50]
		MTBLS656	LTMB0079 [51]
		MTBLS524	LTMB0080 [52]
		MTBLS100	LTMB0081 [53]
	Breath	MTBLS6760	LTMB0085 [54]

### 3.3 Proteïnes i gens

Pel que fa a les proteïnes, primer incorporem els compostos disponibles a les bases de dades públiques, després afegim la literatura rellevant. Tot i revisar els repositoris més usats en proteòmica (MassIVE [55] i PRIDE [56]), no hi hem trobat estudis específics per a saliva i alè. Els gens van associats a les proteïnes i ja s'obtenen amb les fonts esmentades.

**Base de dades de proteïnes:** Les cinc bases de dades utilitzades pel que fa les proteïnes, han sigut les següents.

"MarkerDB" o MDB ( <https://markerdb.ca/> ), ja esmentada en l'apartat dels metabòlits.

La "ExoCarta" o EXOC ( <http://www.exocarta.org/> ) és una plataforma en línia on es recopila i organitza informació sobre proteïnes, ARN i lípids provinents de nombrosos estudis exosomals. Aquesta informació està classificada segons els estàndards mínims establerts per la Societat Internacional de Vesícules Extracel·lulars per garantir una definició precisa de les vesícules extracel·lulars [57].

La "Vesiclepedia" o VSCP ( <http://www.microvesicles.org/> ) és una plataforma en línia que recopila informació sobre ARN, proteïnes, lípids i metabòlits presents a les vesícules extracel·lulars. Aquesta base de dades inclou dades provinents d'una àmplia varietat d'estudis, publicats i no publicats. Actualment, Vesiclepedia acull una gran quantitat d'informació, amb nombroses entrades d'ARN, proteïnes i lípids/metabòlits recopilats de diferents investigacions sobre vesícules extracel·lulars[58].

La "Human Body Fluid Proteome" o HBFP ( <https://bmbl.bmi.osumc.edu/HBFP/> ) es centra en l'estudi del proteoma validat experimentalment en 17 tipus diferents de fluids corporals humans [59]. És la única base de dades de la qual s'ha pogut extreure dades dels dos biofluids.

La "Human Salivary Proteome" o HSP ( <https://salivaryproteome.org/> ) és una plataforma de dades públiques que permet investigar i accedir a informació específica sobre el proteoma de la saliva. La HSP s'actualitza de forma dinàmica utilitzant estudis publicats sobre el proteoma de la saliva i els registres de referència actualitzats de proteïnes. Aquesta plataforma integra informació de diverses fonts, incloent-hi bases de dades externes de proteïnes, genomes, transcriptomes i glicoma, per proporcionar una visió completa i actualitzada del proteoma de la saliva[60].

**Literatura de proteïnes:** La cerca de bibliografia científica s'ha realitzat de manera similar al cas dels metabòlits, però per a proteïnes. Obtenint un únic article que incorporés una base de dades: Lacombe et al., 2018[30], el qual és un estudi que treballa la caracterització proteòmica del condensat de la respiració exhalada humana, per tant la major part de proteïnes trobades van relacionades amb l'alè, encara que s'han trobat unes poques amb la saliva.

### 3.4 Elements químics

Pels elements químics, s'han cercat en els tres tipus de fonts.

#### **Base de dades dels elements químics:**

La "Human Metabolome Database" o HMDB ( <https://hmdb.ca/> ), on els elements químics es trobaven juntament amb els metabòlits. Aquests s'ha filtrat i separat. Únicament contenia elements químics trobats a la saliva.

La "MarkerDB" o MDB ( <https://markerdb.ca/> ). Aquesta base de dades també conté únicament elements químics trobats a la saliva.

#### **Literatura dels elements químics:**

Un article amb base de dades contenia elements químics tant de la saliva com de l'alè: LTMB0001: Drabinska et al., 2021[20].

#### **Repositoris dels elements químics:**

El "Metabolomics Workbench" ( <https://www.metabolomicsworkbench.org/> ), ha retornat dos estudis per saliva (LTMB0071 i LTMB0075), i un per alè (LTMB0084). A les Taules 5 i 6, estan descrits a quins estudis corresponen, respectivament.

## 4 Noves bases de dades de la composició de la saliva i l'alè

### 4.1 Creació de la base de dades

Cada font d'informació òmica sobre la composició dels dos biofluids presenta les seves pròpies bases de dades o resultats de manera individual.

En el cas dels metabòlits, seguim una estratègia on s'assigna a cada compost una identificació única, l'InChIKey[61]. Creem noves bases de dades de la saliva i l'alè combinant aquestes identificacions InChIKey.

Per a les proteïnes, també seguim una estratègia similar. Utilitzem un identificador únic anomenat Uniprot per a cada proteïna i els combinats per construir la base de dades.

En quant els elements químics, donat que també disposen de InChIKey, apliquem la mateixa estratègia que per als metabòlits.

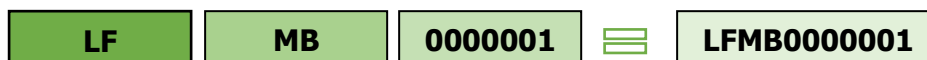
### 4.2 Atributs de la base de dades

La nova base de dades de la saliva i l'alè combina metabòlits, proteïnes i elements químics, cadascun amb els seus propis atributs depenent del tipus de molècula que estiguem analitzant. L'estratègia utilitzada per classificar aquests components variarà en funció de les característiques específiques.

#### Atributs dels metabòlits:

Després de revisar la informació disponible sobre metabòlits i considerar les necessitats dels futurs usuaris de la base de dades, es van triar un total de 18 característiques rellevants.

- *LifeFluids ID*: Una codificació comuna a totes les bases de dades de LifeFluids. El seu objectiu és classificar totes les molècules d'aquest projecte utilitzant una nomenclatura única. Està compost per dues abreviatures, LF (Life Fluids) i MB (metabòlits), seguides de 7 caràcters que descriuen la posició on es troba el metabòlit. A continuació es mostra un exemple del primer identificador.



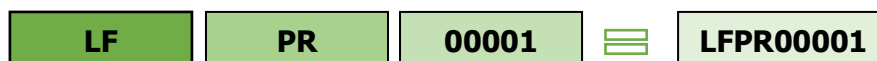
- *RefMet*: És una forma de nomenar de manera estandarditzada tant les estructures de metabòlits individuals com les espècies de metabòlits identificades mitjançant tècniques espectroscòpiques en experiments de metabolòmica.
- *Massa molecular*: S'obté calculant la suma dels pesos atòmics dels àtoms que formen la fórmula molecular de la substància[62].
- *Fórmula química*: És una representació escrita dels elements químics presents en un compost, indicant la proporció en què són presents[63].
- *InChIKey*: "International Chemical Identifier Key" és un codi alfanumèric únic que s'utilitza per identificar de manera precisa i unívoca un compost químic. Va ser creat per l'Institut Nacional d'Estàndards i Tecnologia dels Estats Units (NIST) com a part de l'estàndard InChI[64].
- *HMDB ID*: és la codificació utilitzada a la base de dades HMDB[33]. HMDB és la base de dades més rellevant sobre metabòlits trobats al cos humà.
- *PubChem ID*: Conegut com a PubChem CID, és l'identificador utilitzat a PubChem[65], la base de dades química desenvolupada per l'Institut Nacional de Salut [66] (NIH, per les sigles en anglès) de la Biblioteca Nacional de Medicina.
- *KEGG ID*: És l'identificador utilitzat a KEGG (Enciclopèdia de gens i genomes de Kyoto)[18].

- *BinVestigate ID*: És l'entrada a l'aplicació web de BinVestigate per a compostos obtinguts mitjançant GC-MS. Els bins estan mínimament definits per les següents propietats: espectre de masses, índex de retenció (RI), massa de quantificació, llista de masses úniques i un nombre d'identificació únic[67].
- *Lipid Maps ID*: És una nomenclatura utilitzada al servei web de Lipid Maps, una base de dades restringida per a informació sobre lípids. Lipid Maps proporciona accés a la nomenclatura de lípids, bases de dades, eines, protocols, estàndards, tutorials, reunions, publicacions i altres recursos[68].
- *Drugbank ID*: És una nomenclatura utilitzada per identificar els compostos de la base de dades de Drugbank. Drugbank és una base de dades integral i d'accés gratuït que conté informació sobre medicaments i objectius terapèutics[69].
- *Ontologia Química*: Classificacions químiques dels compostos.
- *Referències*: Font original (bases de dades, literatura o repositoris).

### Atributs de les proteïnes:

Després de revisar la informació disponible sobre les proteïnes i avaluar les necessitats dels futurs usuaris de la base de dades, es van seleccionar un total de 5 atributs.

- *LifeFluids ID*: Una codificació comuna a totes les bases de dades de LifeFluids. El seu objectiu és classificar totes les molècules d'aquest projecte utilitzant una nomenclatura única. Està compost per dues abreviatures, LF (Life Fluids) i PR (proteïnes), seguides de 5 caràcters que descriuen la posició on es troba el metabòlit. A continuació es mostra un exemple del primer identificador.

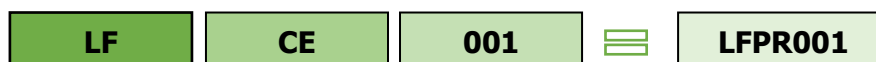


- *Nom*: Nom de la proteïna.
- *Gene Symbol*: Símbol de codificació genètica de la proteïna. El flux d'informació de l'ADN a les proteïnes és un dels principis fonamentals de la biologia molecular[70]. Per aquesta raó, volem marcar la codificació genètica subjacent.
- *UniProt ID*: És un identificador únic assignat a una seqüència de proteïna específica en una base de dades. L'UniProt d'una proteïna és essencial per fer cerques i referenciar informació sobre aquesta proteïna en estudis científics i bases de dades bioinformàtiques.
- *Referències*: Font original (bases de dades, literatura o repositoris).

### Atributs dels elements químics:

En el cas dels elements químics els atributs són similar als casos anteriors.

- *LifeFluids ID*: La codificació és comuna a totes les bases de dades de LifeFluids. Està composta per dues abreviatures, LF (Life Fluids) i CE (elements químics, en anglès: *Chemical Elements*), seguides de 3 caràcters que descriuen la posició en què es troba l'element químic a la taula periòdica. A continuació es mostra un exemple d'aquesta codificació amb l'element numero 1 de la taula periòdica (hidrogen).



- *Nom*: Nom de l'element químic.
- *Símbol*: Codificació que s'utilitza en la taula periòdica.
- *HMDB ID*: és la codificació utilitzada a la base de dades HMDB[33].
- *PubChem ID*: Conegut com a PubChem CID, és l'identificador utilitzat a PubChem[65].



- *Classificació de la Taula Periòdica dels Elements*: Els metalls i altres elements químics es classifiquen segons el seu grup principal (metalls, metal·loides i no metalls), i després per famílies, que són grups d'elements amb propietats similars.
- *Referències*: Font original (bases de dades, literatura o repositoris).

### 4.3 Recull d'atributs

Cada compost, proteïna i element químic, pot estar recollit en les diverses bases de dades utilitzades amb descriptors i atributs diferents. Per tal d'homogeneïtzar els resultats, fem servir diferents eines que ens permeten obtenir els atributs prèviament descrits a la secció 4. Aquestes eines funcionen com a motors, on ingresem un atribut i obtenim un o més d'un atribut com a resultat.

En el cas dels **metabòlits**, utilitzem l'InChIKey com a component principal de la base de dades. Aquest identificador és únic per a cada compost i ens permet connectar els atributs entre aquests (Figura 2). Si haguéssim fet servir noms en lloc d'identificadors, hauria estat més difícil obtenir la resta de la informació a causa de les variacions de noms entre les bases de dades.

- Els noms i l'identificador PubChem es van obtenir a través del servei d'intercanvi d'identificadors de PubChem (<https://pubchem.ncbi.nlm.nih.gov/identifiers/idexchange.cgi>). Aquest servei permet establir connexions entre diferents atributs com InChIKey, SMILES, sinònims i identificador PubChem.
- El pes molecular, la fórmula, l'identificador KEGG, l'identificador de Lipid Maps i l'identificador Drugbank s'obtenen mitjançant el servei de traducció química (CTS)[71]. Aquest motor converteix certs tipus d'entrades (InChIKey, noms químics, etc.) a diferents sortides, com les esmentades anteriorment.
- Els atributs de superclasse, classe i subclasse s'obtenen a partir de la "ClassyFire Batch Compound Classification"[72] És una aplicació basada en web per a la classificació estructural automatitzada de les entitats químiques.
- El Bin ID és proporcionat per l'aplicació web BinVestigate[19], utilitzada per explorar els compostos trobats a la Base de dades BinBase. L'atribut d'entrada necessari per obtenir el bin ID és InChIKey.
- L'identificador de HMDB és proporcionat pel servei web "Human Metabolome Database"[16], tanmateix també es pot extreure amb el CTS[71].
- InChIKey, sinònims i identificador de PubChem.
- El RefMet s'obté a partir del repositori Metabolomic Workbench[25]

L'esquema d'interconnexió d'atributs dels metabòlits (Figura 2), guiat per l'InChIKey, proporciona un marc per a la recopilació, l'anàlisi i l'exploració exhaustiva de dades relacionades amb els metabòlits.

En el cas de les **proteïnes**, els atributs Uniprot i Gene symbol estaven disponibles a cada literatura consultada, per la qual cosa no va ser necessari utilitzar cap eina de cerca per obtenir-los. No obstant això, també es poden trobar a la base de dades de Uniprot (<https://www.uniprot.org/>). En el cas que s'hagués de dur a terme una recerca amb una gran quantitat de proteïnes per buscar un d'aquests atributs, també es pot utilitzar la pàgina web de bioDBnet ( <https://biodbnet-abcc.ncifcrf.gov/> ).

En el cas dels **elements químics** l'únic atribut que s'ha hagut de buscar si ha fet falta, ha sigut el símbol i s'ha trobat en la taula periòdica de la Royal Society of Chemistry[73].

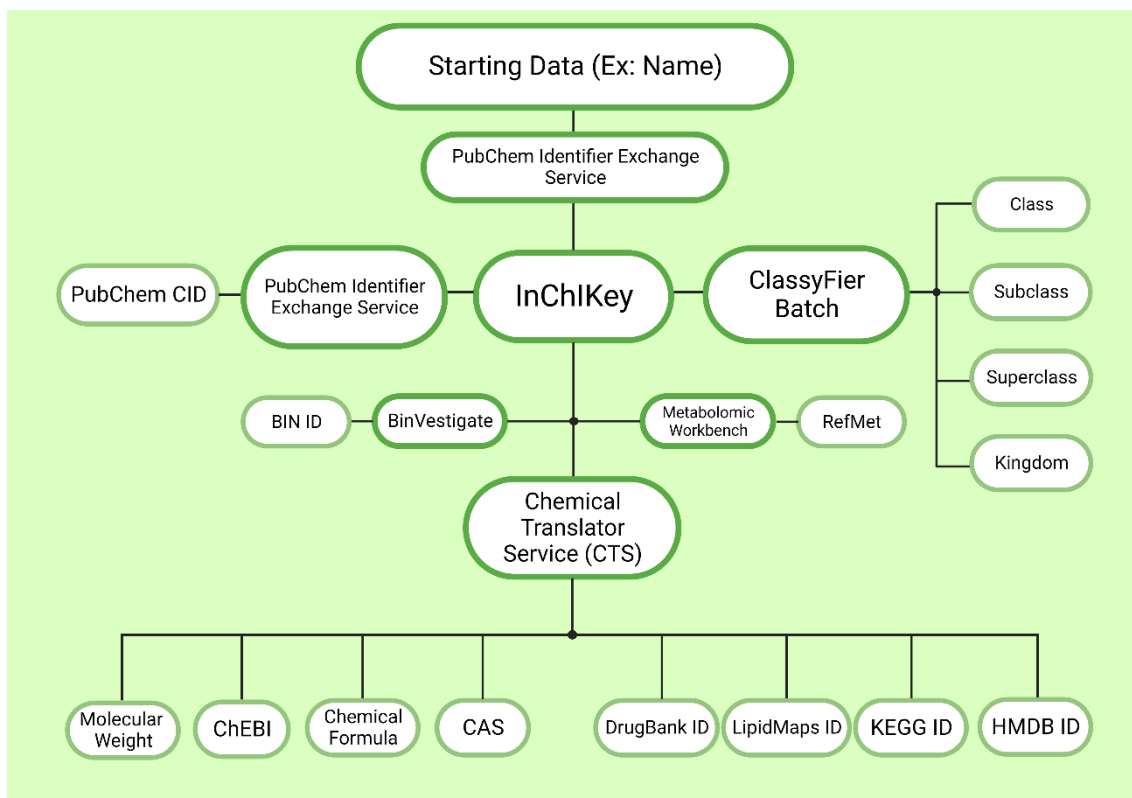


Figura 2. Descripció de l'obtenció dels principals atributs de la base de dades de metabòlits.

#### 4.4 Desenvolupament del codi

El programari utilitzat per crear el codi s'anomena Anaconda[83]. És un entorn de desenvolupament que facilita la programació en els llenguatges Python i R, especialment per a aplicacions científiques. Utilitzem Python per combinar els fitxers de les bases de dades, literatura i repositoris, per tal de crear una nova base de dades que inclogui els atributs esmentats anteriorment. L'estratègia seguida es mostra a la Figura 3.



Figura 3. Estratègia per desenvolupar el codi.



### Base de dades dels metabòlits:

Per començar, es van combinar tots els compostos trobats en les principals bases de dades, [19] descrits en la secció 3.2. L'objectiu principal del codi era poder inspeccionar cada base de dades, i afegir aquells metabòlits que no hi fossin presents. La principal eina de guia entre les diverses fonts i la nostra base de dades ha sigut la InChIKey, tal i com s'especifica a la secció 4.1. A l'hora de fer servir el codi, ens podem trobar dues opcions: primera, que el compost ja aparegués en la nostra base de dades i en aquest cas s'afegiria la seva procedència a l'apartat de referències (Figura 4a), o el cas de que no existís a la nostra base de dades i en aquest cas es van anar afegint a partir de la última posició (Figura 4b). Una vegada que afegits a la base de dades creada, es van buscar els atributs que faltessin, com s'ha descrit a la secció 4.3.

A) Cas 1 – Entrada ja existent			B) Cas 2 – Entrada nova		
	Inchikey	DB		Inchikey	DB
		HMDB   BV			
					BV

Figura 4. Creació de la base de dades: (A) Cas 1, on ja existeix el compost i s'afegeix la seva procedència a referències. (B) Cas 2, no existeix el compost i s'afegeix al final.

### Base de dades de proteïnes:

Com en els metabòlits, per començar, es van afegir les proteïnes que ens proporcionaven les bases de dades que podem trobar a la Taula 2 per la saliva i a la Taula 3 per l'alè. Posteriorment es van afegir les proteïnes trobades als articles. El procediment va ser molt semblant al que es va fer servir per al metabòlits, en el cas de les proteïnes la principal eina de guia ha sigut el UniProt, tal i com s'especifica a la secció 4.1.

### Base de dades d'elements químics:

Com en els metabòlits, hem fet servir la InChIKey per a combinar els elements químics literatura disponibles en les diferents fonts utilitzades, tal com es mostra a la Taula 2 per la saliva i a la Taula 3 per l'alè.

### Curació manual

Cada base dada de cada biofluid, se li ha realitzat una curació manual per assegurar-nos eliminar duplicats, dades incorrectes i completar correctament els atributs.

## 5 Base de dades Life Fluids de la Saliva

### 5.1 Life Fluids SALIVA metabòlits

En metabòlits, primer afegim les tres **bases de dades** principals. L'HMDB serà la nostra "Gold Standard". En l'àmbit de la ciència de dades, "Gold Standard" fa referència a un conjunt de dades de referència que s'utilitza per comparar i validar els resultats[74]. En l'HMDB es van detectar 1.239 metabòlits, però, només se'n van incloure 1.195 després d'eliminar duplicats i els elements químics. A la "BinVestigate", es van detectar 4.871 "bins" en mostres de saliva, dels quals 3.922 van ser descartats perquè eren compostos desconeguts i no tenien un InChIKey associat o eren estàndards de referència. Dels 949 "bins" restants, només se'n van incloure 497 després d'eliminar els duplicats, i es van afegir un total de 297 nous InChIKey com a nous metabòlits. De la "MarkerDB", es van incorporar un total de 27 metabòlits, on tots van coincidir amb els compostos ja existents.

Pel que fa a la **literatura**, a Drabinska et al., 2021[21], es van poder extreure un conjunt de dades de 857 metabòlits. Després d'una curació manual i eliminar duplicats, es van incorporar un total de 468 metabòlits, 177 nous compostos respecte els que ja ens proporcionava l'HMDB. A Pereira et al., 2020[22], comencem amb un conjunt de dades de 42 metabòlits. Després d'una curació manual, es van incloure un total de 36 metabòlits, 7 nous respecte l'HMDB. A Gardner et al., 2020[23], es va aconseguir 343 metabòlits. Després de la curació manual, es van incloure 104 metabòlits, dels quals 30 van ser noves entrades a la nostra base de dades. Huang et al., 2023[24] comencem amb un conjunt de dades de 40 metabòlits. Després de la curació manual i treure els duplicats, es van incloure un total de 33 metabòlits, 8 nous respecte a l'HMDB.

Al nostre treball de recerca, ens hem centrat en dos **repositoris** principals de dades, Metabolomic Workbench i MetaboLights.

En Metabolomic Workbench[25], trobem 7 estudis que se centraven específicament a la saliva. Tot i això, només 6 d'aquests estudis tenien informació útil. En vam poder extreure un conjunt de dades de 887 metabòlits. Un cop eliminats els duplicats i realitzada una curiosa curació de les dades, ens quedem amb 518 metabòlits únics. 149 d'aquests metabòlits, nous respecte l'HMDB.

Per altra banda, explorem MetaboLights[26]. Aquí també trobem 7 estudis sobre la saliva, però igual que en el cas anterior, només 6 oferien dades útils per a la nostra feina. D'aquests estudis, en vam poder extreure un total de 747 metabòlits. Després del procés d'eliminació de duplicats i curació manual de les dades, ens vam quedar amb 425 metabòlits únics.

La Figura 5 ens mostra un resum, on es pot veure del total de la nostra base de dades nova (1.877 metabòlits) el percentatge de noves entrades (36%) i el percentatge de metabòlits que ja existien a la nostra "Gold Standard" (64%).

La Figura 6 ens mostra les dades que ha adquirit la Life Fluids Saliva respecte la nostra base de dades de referència.

En resum, **la base de dades Life Fluids de Saliva per a metabòlits està composta per 1.877 metabòlits**, dels quals 1.492 provenen de bases de dades, 222 de la literatura consultada i 163 de repositoris.

### LIFE FLUIDS SALIVA - METABÒLITS

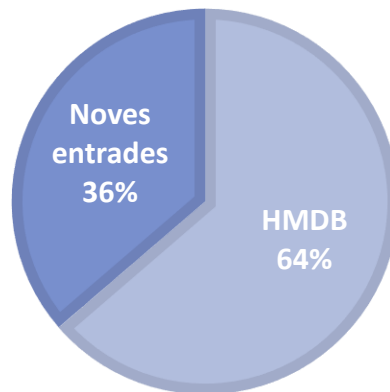


Figura 5. Metabòlits de la base de dades Life Fluid Saliva, segons si ja estaven recollits a la HMDB o si són noves entrades.

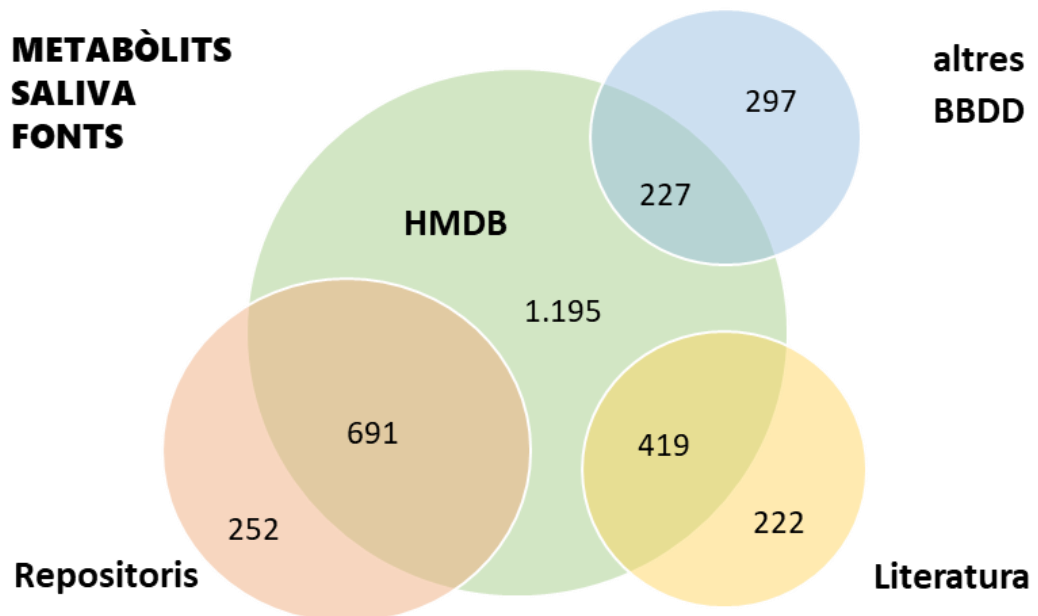


Figura 6. Base de dades metabòlits Life Fluids Saliva segons les fonts consultades.

## 5.2 Life Fluids SALIVA proteïnes

En **proteïnes**, la nostra "Gold Standard", "Human Salivary Proteome" (HSP), originalment va contenir 3.457 proteïnes. La tasca principal aquí va ser la curació de dades. Com a resultat d'aquesta curació, ens vam quedar amb 3.442 proteïnes.

Com s'ha detallat a l'apartat 3.3, també hem analitzat altres bases de dades. Una d'aquestes és l'"Exocarta", que inicialment va mostrar 458 proteïnes. Novament, després de la curació de dades i l'eliminació de duplicats, la xifra es va reduir a 431 proteïnes. D'aquestes, 129 eren noves entrades respecte a l'HSP. A més, hem fet servir la base de dades de la "Vesiclepedia", que inicialment va presentar 655 proteïnes. Posteriorment, amb l'aplicació de la curació de dades, la xifra es va ajustar a 623 proteïnes. Aquí, vam descobrir 219 entrades noves que no eren presents a l'HSP.

Finalment, hem explorat la base de dades de l'"Human Body Fluid Proteome" (HBFP), que contenia inicialment 2.873 proteïnes. Després de la curació, ens vam quedar amb 2.722 proteïnes, incloent 1.254 noves entrades respecte a l'HSP.

No obstant això, cal destacar que l'única font bibliogràfica que ens ha proporcionat informació útil sobre les proteïnes de la saliva és l'article de Lacombe et al., 2018[30], que esmenta un total de 8 proteïnes. Totes aquestes proteïnes ja es trobaven presents a l'HSP.

La Figura 7 ens mostra un resum, on es pot veure del total de la nostra base de dades nova (3442 proteïnes) el percentatge de noves entrades (36%) i el percentatge de metabòlits que ja existien a la nostra "Gold Standard" (64%).

La Figura 8 ens mostra les dades que ha adquirit la Life Fluids Saliva respecte la nostra base de dades de referència.

En resum, **la base de dades Life Fluids de Saliva per a proteïnes està composta per 4.810 proteïnes**, dels quals 3.422 provenen de la Human Salivary Proteom (HSP) i 1.368 provenen d'altres bases de dades, tal i com es mostra en la Figura 8.

### LIFE FLUIDS SALIVA - PROTEÏNES

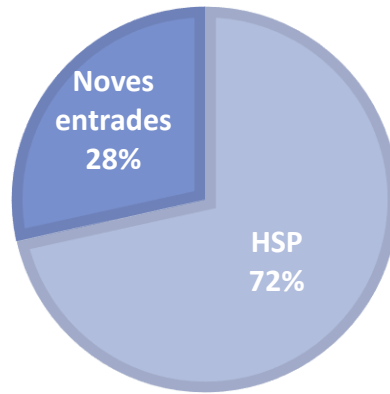


Figura 7. Proteïnes de la base de dades Life Fluid Saliva, segons si ja estaven recollides a la HSP o si són noves entrades.

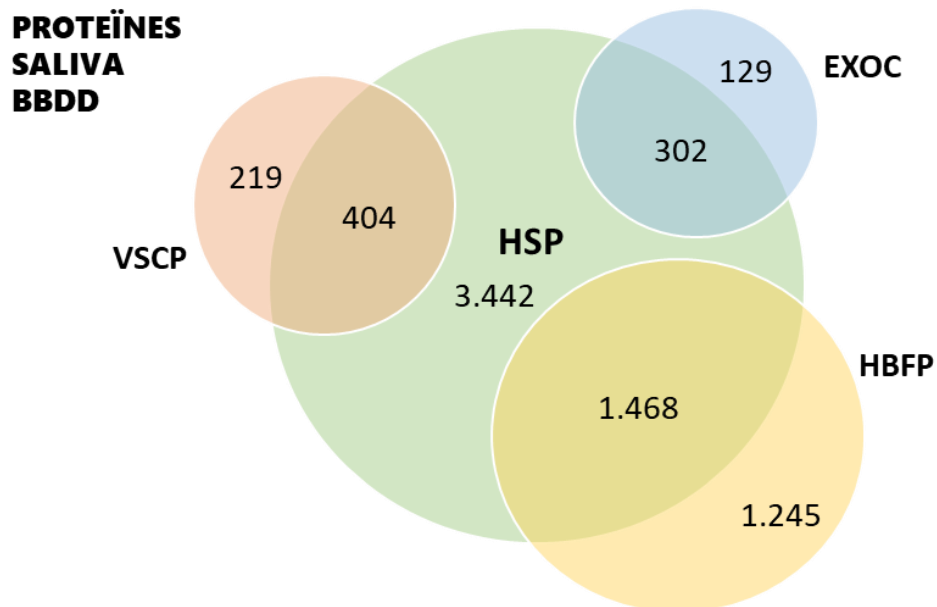


Figura 8. Esquema de la base de dades de proteïnes de Life Fluids Saliva segons la base de dades consultada.

### 5.3 Life Fluids SALIVA elements químics

Dins dels **elements químics** identificats, s'ha pogut registrar un total de 40 elements. Aquests han estat extrets inicialment de l'HMDB.

Un cop identificats, es va procedir a filtrar aquests elements amb l'objectiu de refinar la informació i eliminar possibles errors. D'aquesta manera, es va començar a construir una base de dades específica, la qual partia dels 40 components detectats a l'HMDB. Subratllar que, dins d'aquest conjunt, solament un element, el Ferro, es va poder identificar en una altra base de dades, en aquest cas MarkerDB.

En la revisió de la literatura científica, l'article de Drabinska et al., 2021[21] va resultar particularment útil. En aquest estudi, es va poder extreure informació relativa a 22 elements químics. Cal destacar que tots aquests elements ja es trobaven en la nostra base de dades.

Pel que fa als repositoris de dades, la nostra investigació a Metabolomic Workbench[25] va permetre la detecció de 2 elements químics. Ambdós, igual que en el cas de la literatura científica, ja es trobaven en la nostra base de dades.

La Figura 9 ens mostra les dades que ha adquirit la Life Fluids Saliva respecte la nostra base de dades de referència.

En resum, **la base de dades Life Fluids de Saliva per a elements químics està composta per 40 elements**, dels quals ningun prové de noves fonts. Tot els elements químics afegits ja existien prèviament respecte la nostra base de dades de referència (Figura 9).

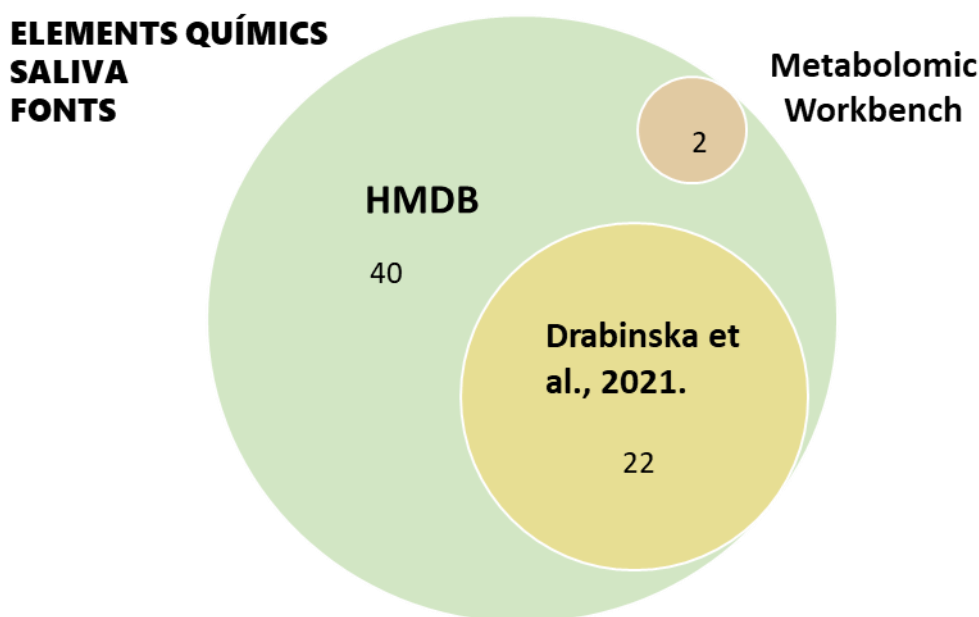


Figura 9. Esquema de la base de dades d'elements químics de Life Fluids de la Saliva. Ens mostra les dades adherides i ja existents provinents de las diferents fonts.

## 6 Base de dades Life Fluids de l'alè

### 6.1 Life Fluids ALÈ metabòlits

Per a l'anàlisi dels metabòlits a l'alè, seguim un procediment similar al que fem servir per saliva. Primer, vam mirar les **bases de dades**, essent de nou la nostra "Gold Standard", la base de dades HMDB, de la qual vam obtenir 338 metabòlits. Aquesta base constituïria el punt de partida per al nostre treball posterior.

Utilitzem també la base de dades "Knapsack", que ens va proporcionar un total de 29 metabòlits. 3 d'aquests metabòlits no es trobaven a la base de dades HMDB.

La **literatura** científica va ser una altra font d'informació significativa per a nosaltres. D'un estudi específic, Drabinska et al., 2021[21], vam poder extreure'n un total de 1.473 compostos. Després del procés de curació de dades, ens vam quedar amb 1.069 metabòlits, que van afegir 1.023 noves entrades a la nostra base de dades. Un altre article rellevant va ser el de Sharma et al., 2023[32], que ens va proporcionar inicialment 313 metabòlits. Després de la curació i eliminació de duplicats, ens vam quedar amb 168, dels quals 35 eren nous respecte a la nostra base de dades inicial.

Pel que fa als **repositoris** de dades, a Metabolomic Workbench trobem 3 estudis relacionats amb l'alè. No obstant això, un no oferia informació útil, per la qual cosa va ser descartat. Dels dos estudis restants, vam poder obtenir inicialment 155 components. Després del procés de curació manual i eliminació de duplicats, van quedar 107 metabòlits útils, 17 dels quals eren nous respecte a la nostra base de dades de referència. Per últim, recorrem al repositori MetaboLights, on hi trobem 4 estudis relacionats amb el nostre biofluid d'interès. Tanmateix, només un proporcionava una llista de metabòlits. D'aquest estudi, vam obtenir 5 metabòlits, tots nous per a la nostra base de dades.

La Figura 10 ens mostra del total de la nostra base de dades nova (1415 metabòlits) el percentatge de noves entrades (76%) i el percentatge de metabòlits que ja existien a la nostra "Gold Standard" (24%).

La Figura 11 ens mostra les dades que ha adquirit la Life Fluids de l'alè respecte la nostra base de dades de referència.

En resum, la **base de dades Life Fluids de l'Alè per a metabòlits està composta per 1.465 metabòlits**, dels quals 391 provenen de bases de dades, 1.057 de la literatura consultada i 17 de repositoris.

### LIFE FLUIDS ALÈ - METABÒLITS

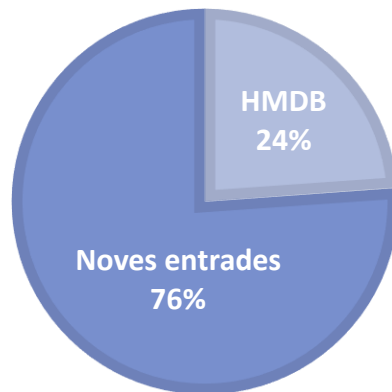


Figura 10. Metabòlits de la base de dades Life Fluid Alè, segons si ja estaven recollits a la HMDB o si són noves entrades.

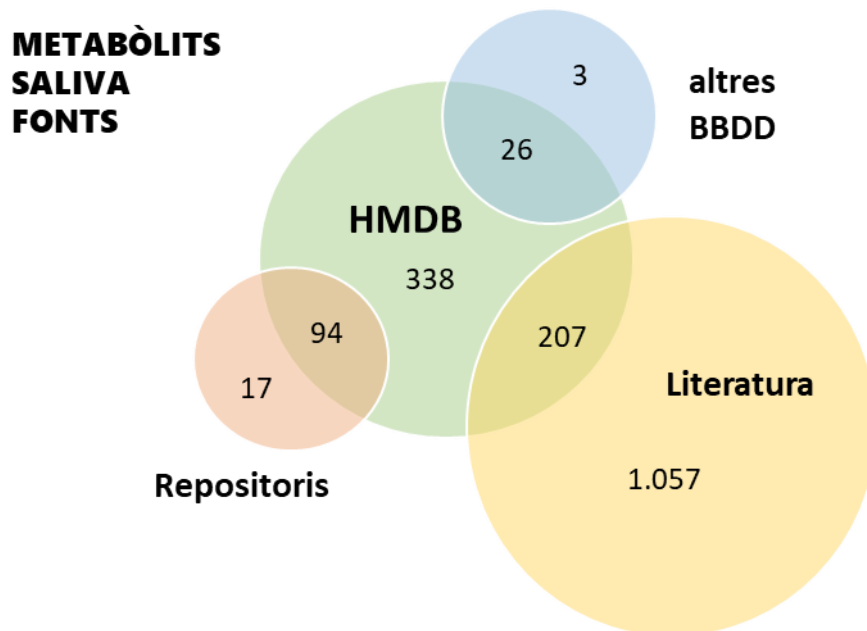


Figura 11. Base de dades metabòlits Life Fluids de l'alè, segons les fonts consultades.



## 6.2 Life Fluids ALÈ proteïnes

Al nostre estudi de les **proteïnes** a l'alè, l'única base de dades que vam poder trobar amb informació rellevant i que també es va convertir en la nostra "Gold Standard", com en el cas de la saliva, va ser el "Human Body Fluid Proteome" (HBFP). Ressaltar que escollim aquesta base de dades de referència perquè a diferència de la saliva, en aquest biofluid no hi ha una base de dades només de l'alè com podia ser la HSP. Aquesta base de dades inicialment contenia 358 proteïnes. Després del procés de curació i eliminació de duplicats, vam quedar amb 351 proteïnes que van formar la base de la nostra base de dades.

A més de la HBFP, fem una cerca bibliogràfica per trobar més informació. Només trobem un article que era rellevant per a la nostra investigació: Lacombe et al., 2018. Aquest estudi va proporcionar dades sobre 146 proteïnes útils. 4 d'aquestes proteïnes van resultar noves entrades per a la nostra base de dades.

En la següents Figura 12 podem veure el resultats sobre les proteïnes en l'alè. En la Figura 12A tenim els percentatges de les noves entrades (1%) i les dades de la HBFP (99%). Juntament tenim la Figura 12B on es recull el nombre de proteïnes inicials de la nostra base de dades de referència i les que vam trobar a la única literatura literària útil.

La Figura 12B ens mostra les dades que ha adquirit la Life Fluids Saliva respecte la nostra base de dades de referència.

En resum, **la base de dades Life Fluids de alè per a proteïnes està composta per 355 proteïnes**, dels quals 4 provenen d'altres fonts mostrades en la Figura 12 B.

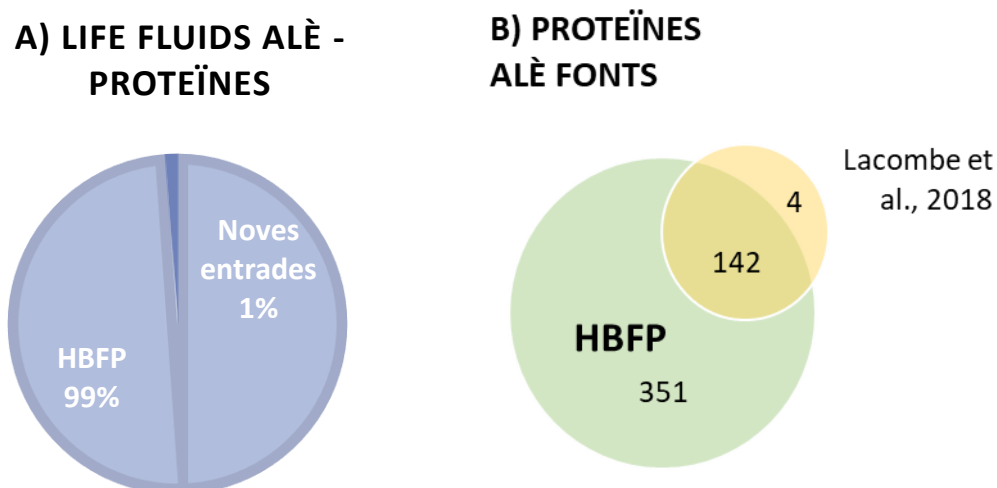


Figura 12. A) Proteïnes de la base de dades Life Fluid Alè, segons si ja estaven recollits a la HMDB o si són noves entrades. B) Esquema de la base de dades de proteïnes de Life Fluids Alè segons les fonts consultades.

### **6.3 Life Fluids ALÈ elements químics**

En referència als **elements químics** presents a l'alè, no se n'han trobat en les bases de dades.

En la revisió de la literatura científica, en l'article de Drabinska et al., 2021[21] se n'han pogut incorporar 2.

Pel que fa als repositoris de dades, la nostra investigació a Metabolomic Workbench[25] va permetre la detecció de 1 altre elements químics no inclòs previament.

En resum, **la base de dades Life Fluids de Saliva per a elements químics està composta per 3 elements**, dels quals sol 1 prové de un font del repositori Metabolomic Workbench.

## 7 Correlacions entre les dades Life Fluids de saliva i d'alè

La relació entre la saliva i l'alè és important sobretot en el context de contaminació del tràfic ja que comparteixen conductes i òrgans per on s'extreuen els biofluids [75]. Per això, vam creure important poder fer una llista de metabòlits i proteïnes en comú al nostre estudi.

### Correlació entre metabòlits

Pel que fa en **metabòlits**, de les taules ja existents amb els compostos trobats, agrupats i ordenats anteriorment, un total de 1.877 metabòlits en la saliva i 1.415 en l'alè, s'han pogut identificar un total de **518 compostos en comú**. La Figura 13A ens mostra la interacció de metabòlits entre els dos biofluids estudiats.

### Correlació entre proteïnes

En el camp de les **proteïnes** hem pogut identificar un total de 4.834 proteïnes a la saliva i 355 pel que es refereix a l'alè. Pel que fa la relació entre les dos taules, han sigut un total de **321 compostos en comú**. La Figura 13B en mostra la interacció de proteïnes entre els dos biofluids estudiats.

### Correlació entre elements químics

No s'ha trobat cap element químic comú entre ambdós biofluids, tal i com es mostra a la Figura 13C.

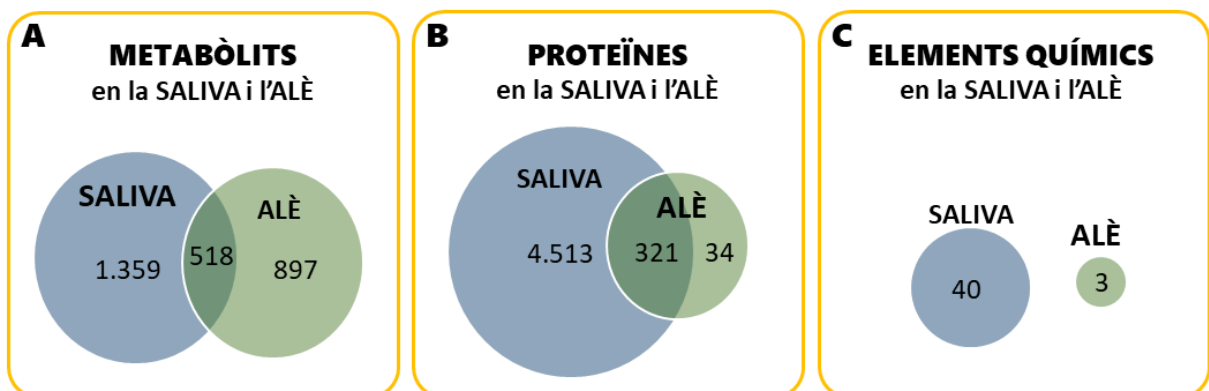


Figura 13. Gràfica que ens mostra el nombre comú entre els dos biofluids per a: A) metabòlits; B) proteïnes; i C) elements químics.

## 8 Discussió

### 8.1 Creació, atributs i curació

S'han creat dues bases de dades que comparteixen, en part, bio-localització com són la saliva i l'alè. Per tal d'obtenir unes taules de la composició d'aquests biofluids el més fidedigna possible, s'han recopilat, curat i ajuntat dades de bases de dades, repositoris i articles científics (on proveïssin una base de dades). S'obté d'aquesta manera dues bases de dades: 1) Life Fluids Saliva amb 1.877 metabòlits, 4.810 proteïnes, i 40 elements químics; i 2) Life Fluids Alè amb 1.467 metabòlits, 355 proteïnes, i 3 elements químics.

Tenir un conjunt divers d'atributs en una base de dades de **metabòlits** proporciona un panorama més complet i permet una anàlisi més profunda de les dades. Cada atribut proporciona un tipus d'informació diferent i pot ajudar a respondre diferents tipus de preguntes. La **massa molecular** i la **formula química**, són útils per a diverses tècniques analítiques i entendre la composició del metabòlit. **HMDB ID, PubChem ID, KEGG ID, BinVestigate ID, Lipid Maps ID i Drugbank** són identificadors de bases de dades que permeten la vinculació i l'encreuament de referències amb diverses bases de dades rellevants. Això permet agregar informació i ajuda a integrar dades de múltiples fonts per a una anàlisi més profunda. Destacar la dificultat que s'ha tingut per poder aconseguir un gran nombre d'atributs per tal de completar la taula. Per exemple, en la Lipid Maps[76] i la DrugBank[77], que són unes bases de dades que proporcionen informació detallada sobre els lípids i les drogues respectivament i les seves vies metabòliques, en la nostra taula de la saliva tenim un **35,42%** pel que fa a Lipid Maps ID, **14,38%** en Drugbank ID i en l'alè un **11,94%** i **8,74%** respectivament. Això pot ser degut a que la saliva i l'alè poden no contenir nivells detectables de certs metabòlits degut a que les tècniques actuals per analitzar els metabòlits als biofluids poden no ser prou sensibles, en comparació amb altres biofluids com l'orina o la sang.

Destacar la importància de la curació manual i l'eliminació dels duplicats. **La curació manual implica la revisió, verificació i correcció de les dades a la base de dades.** Aquest procés pot implicar verificar l'exactitud de les dades, corregir errors, omplir camps que falten i assegurar que les dades siguin coherents i estiguin en el format correcte. La curació manual és especialment important en camps com la biomedicina, on els errors o inconsistències a les dades poden tenir conseqüències serioses. A més, la curació manual permet incorporar coneixement expert a la base de dades, cosa que pot millorar enormement la seva qualitat i utilitzada.

La eliminació de duplicats també té una gran rellevància sobretot en aquest tipus de treballs on es recopila informació de múltiples fonts. Aquests duplicats poden ser deguts a diverses raons, incloent variacions en la nomenclatura, errors d'entrada o la inclusió de la mateixa informació a partir de diferents fonts. Els duplicats poden causar una sèrie de problemes, incloent-hi una sobreestimació de la mida de la base de dades, distorsions en l'anàlisi de les dades i confusions en tractar d'interpretar els resultats. **Eliminar aquests duplicats millora la precisió i la utilitat de la base de dades.**

## 8.2 Metabòlits

Pel que fa als **metabòlits de la saliva**, la nova base de dades conté un total de **1.877 entrades**. La font de referència més reconeguda és la HMDB, on partiem d'un total de 1.195 metabòlits. L'increment de la base de dades amb fonts addicionals, i diversos estudis de la literatura, suposa un augment del **57%** respecte l'anterior treball. (Els càlculs es troben a l'Annex).

La bases de dades s'han ordenat en funció del nombre de bases de dades d'on han sigut afegits i a la vegada també el nombre de fonts literàries on han sigut trobats. Les primeres línies de la base de dades es pot veure a la Figura 14.

NumberID	Nom	RefMet	pted acro	sa	Molecula	Quir	Inchikey	HMDB_ID	Chem	KEGG_ID	Vestigate	lipid_Map	Drugbank	Kingdom	se_Quim	lasse_Quim	parentLevel	Database	Literature	
1	LTMB00001	Tyrosine	Tyrosine		181.1885	C9H11NO	OUYCCCA	HMDB0000		6057	C00082	[16, 1770, 7406, 233]	DB00135	Organic	ci	Carboxylli	Amino ac	Amino ac	HMDB	B.LTMB0000
2	LTMB00002	Hypoxant	Hypoxanthine		136.1115	C5H4N4O	FDGQSTZ	HMDB0000		790	C00262	[1663, 34194]	DB04076	Organic	ci	Imidazop	Purines a	Purinone	HMDB	B.LTMB0000
3	LTMB00003	Phenylala	Phenylalanine		165.1891	C9H11NO	COLNVL	HMDB0000		6140	C00079	[33, 112, 59034]	DB00120	Organic	ci	Carboxylli	Amino ac	Amino ac	HMDB	B.LTMB0000
4	LTMB00004	Proline	Proline		115.1305	C5H9NO2	ONIBWKK	HMDB0000		145742	C00148	[8, 171970, 68068, 1]	DB00172	Organic	ci	Carboxylli	Amino ac	Amino ac	HMDB	B.LTMB0000
5	LTMB00005	Valine	Valine		117.1463	C5H11NO	KZSNJWF	HMDB0000		6287	C00183	[3, 585, 378468]	DB00161	Organic	ci	Carboxylli	Amino ac	Amino ac	HMDB	B.LTMB0000
6	LTMB00006	Creatinin	Creatinine		113.1179	C4H7N3O	DDRJAAN	HMDB0000		588	C00791	[31, 34042, 109708]		Organic	ci	Carboxylli	Amino ac	Amino ac	HMDB	B.LTMB0000
7	LTMB00007	Glucose	Glucose		180.0634	C6H12O6	WQZGKKI	HMDB0000		5793	C00031	[5, 22, 71, 76, 17290, 46121, 88]		Organic	ci	Organoox	Carbohyd	Monosac	HMDB	B.LTMB0000
8	LTMB00008	Creatine	Creatine		131.1332	C4H9N3O	CVSVTCO	HMDB0000		586	C00300	[360149]	DB00148	Organic	ci	Carboxylli	Amino ac	Amino ac	HMDB	B.LTMB0000
9	LTMB00009	Glycine	Glycine		75.0666	C2H5NO2	DHMQDQ	HMDB0000		5257127	C00037	[6, 403, 64525, 2038; DB00145]		Organic	ci	Carboxylli	Amino ac	Amino ac	HMDB	B.LTMB0000
10	LTMB00010	Ornithine	Ornithine		132.161	C5H12N2	AHLPHDH	HMDB0000		6262	C00077	[621, 1821, 1964, 46; DB00129]		Organic	ci	Carboxylli	Amino ac	Amino ac	HMDB	B.LTMB0000
11	LTMB00011	Adenine	Adenine		135.1267	C5H5N5	GFFGJBX	HMDB0000		190	C00147	[1764, 16805, 37831; DB00173]		Organic	ci	Imidazop	Purines a	6-aminop	HMDB	B.LTMB0000
12	LTMB00012	sn-Glycer	sn-Glycerophospho		257.223	C8H20NO	SUHOQUV	HMDB0000		657272	D07349			Organic	ci	Glyceropl	Glyceropl	Glyceropl	HMDB	B.LTMB0000
13	LTMB00013	1-Methyl	1-Methylhistidine		169.1811	C7H11N3	BRMWTN	HMDB0000		7020397			DB04151	Organic	ci	Carboxylli	Amino ac	Amino ac	HMDB	B.LTMB0000
14	LTMB00014	Glutamic	Glutamic acid		147.1293	C5H9NO4	WHUUTDI	HMDB0000		33032	C00025	[28, 617, 62245, 165; DB00142]		Organic	ci	Carboxylli	Amino ac	Amino ac	HMDB	B.LTMB0000
15	LTMB00015	Ethanolar	Ethanolamine		61.0831	C2H7NO	HZAXFHJ	HMDB0000		700	C00189	[3212, 45341, 11356; DB03994]		Organic	ci	Organoni	Amines	Alkanolar	HMDB	B.LTMB0000
16	LTMB00016	Histidine	Histidine		155.1546	C6H9N3O	HNDVDQ	HMDB0000		6274	C00135	[150, 6355, 18490, 5; DB00117]		Organic	ci	Carboxylli	Amino ac	Amino ac	HMDB	B.LTMB0000
17	LTMB00017	Uracil	Uracil		112.0868	C4H4N2O	ISAKRJDG	HMDB0000		1174	C00106	[1664, 3023]	DB03419	Organic	ci	Diazines	Pyrimidir	Pyrimido	HMDB	B.LTMB0000
18	LTMB00018	Caffeine	Caffeine		194.1906	C8H10N4	RYVVLZV	HMDB0000		2519	C07481	[11820, 134461]	DB00201	Organic	ci	Imidazop	Purines a	Xanthine	HMDB	B.LTMB0000
19	LTMB00019	5-Amino	5-Aminopentanoic		117.1463	C5H11NO	JJMDCOV	HMDB0000		138	C00431	[1698, 16; LMFA01100040]		Organic	ci	Carboxylli	Amino ac	Amino ac	HMDB	B.LTMB0000

Figura 14. 20 primeres files de la base de dades de metabòlits en la saliva de Life Fluids.

Després d'una ordenació i anàlisi dels resultats presentats a la taula de la Figura 14, un dels compostos que destaca és la Tirosina. Aquesta substància s'identifica en 13 fonts literàries diferents, posant de manifest la significativa rellevància d'aquest metabòlit en relació amb els efectes que pot tenir sobre el sistema cardíac [23].

A més de la Tirosina, són també notables altres quatre metabòlits: la Hipoxantina, la Fenilalanina, la Prolina i la Valina. Aquests quatre compostos s'identifiquen en un total de 11 fonts diferents, una freqüència que també ressalta la seva possible importància. Entre tots aquests, resulta especialment destacable la snGlycero3phosphocholine, ja que és l'única d'aquesta primera línia de compostos que apareix exclusivament en la base de dades de l'HMDB, a diferència dels altres que també es troben en la BinVestigate.

És necessari mencionar especialment la Valina. S'ha descobert que aquesta substància es troba en nivells elevats en una àmplia varietat de condicions de salut que abasten des de la malaltia periodontal fins al càncer oral, el càncer de mama i el càncer de pròstata. Aquesta observació suggereix que la valina i altres metabòlits semblants poden estar reflectint un canvi patològic no específic en el metaboloma salival, en lloc de ser indicadors específics d'una malaltia concreta [23].

Respecte a la Fenilalanina, és important esmentar que es creu que aquesta pot tenir el potencial de contribuir a la distinció entre els pacients amb càncer de pàncrees i els individus sans en els estudis de control [24]. Aquesta dada, si es confirma, potser podria proporcionar una nova via per a la detecció precoç d'aquest tipus de càncer.

En el cas dels **metabòlits de l'alè**, la nova base de dades conté a un total de **1.465 entrades**. Si comparem amb els 338 metabòlits de la font de referència (HMDB), obtenim un increment del **333%**. L'addició de nous metabòlits respecte a HMDB és notable i parla del potencial de trobar biomarcadors no descoberts i comprendre millor els processos metabòlics als fluids humans.

En l'estudi de l'alè i la seva taula de metabòlits, alguns compostos es destaquen per la seva presència en diverses fonts bibliogràfiques. Aquests inclouen l'Àcid acètic, el Fenol, l'Àcid Fenil acètic, l'Hexanal, el 2-Pentilfuran i el Naftalè. Cadascun d'aquests metabòlits apareix referenciat en almenys tres fonts bibliogràfiques diferents.

És interessant observar que l'Hexanal, el 2-Pentilfuran i el Naftalè no són referenciats en la base de dades "Knapsack", a diferència de l'Àcid acètic, el Fenol i l'Àcid Fenil acètic.

L'Hexanal, en particular, ha rebut atenció com a possible biomarcador del càncer. L'Hexanal es va observar en concentracions més elevades a l'alè exhalat de pacients amb càncer de pulmó. Això implica que l'Hexanal, juntament amb altres compostos, podria estar indicant un canvi bioquímic en l'organisme en resposta a la presència del càncer. No obstant això, aquestes dades encara són preliminars i requereixen més investigació i validació per confirmar la utilitat de l'Hexanal com a biomarcador eficaç[32].

Més enllà del càncer de pulmó, l'Hexanal també s'ha detectat a l'alè de pacients amb càncer de mama, tal com va demostrar un estudi de Li et al.[78]. També, el Naftalè es va trobar present en diverses concentracions a l'alè de pacients amb càncer de mama segons un estudi realitzat per Mangler et al.[79]. Això suggereix que l'Hexanal i el Naftalè podrien tenir un valor potencial com a biomarcadors, no només per al càncer de pulmó, sinó també per al càncer de mama.

### 8.3 Proteïnes

En les **proteïnes**, respecte la **saliva**, hem obtingut un total de **4.810 proteïnes**. La base de dades "Human Salivary Proteome" (HSP) s'ha fet servir de referència o "Gold Standard", amb un total de 3.442 proteïnes. Altres bases de dades com Exocarta, Vesiclepedia i Human Body Fluid Proteome (HBFP) han suposat un **40%** d'entrades noves, cosa que suggereix una alta variabilitat i complexitat de les proteïnes en els fluids humans. Les primeres línies de la base de dades es pot veure a la Figura 15.

A	B	C	D	E	F	G
NumberID	Name	Gene_Symbol	Uniprot	Database	Literature	
1	LFPR00001	AMY1A	PODUB6	EXOC	VS LT	PR0001
2	LFPR00002	CST4	P01036	EXOC	VS LT	PR0001
3	LFPR00003	MUC7	Q8TAX7	EXOC	VS LT	PR0001
4	LFPR00004	PIP	P12273	EXOC	VS LT	PR0001
5	LFPR00005	ZG16B	Q96DA0	EXOC	VS LT	PR0001
6	LFPR00006	ABHD14B	Q96IU4	EXOC	VSCP	HSP   HBFP
7	LFPR00007	ACTA1	P68133	EXOC	VSCP	HSP   HBFP
8	LFPR00008	ACTN1	P12814	EXOC	VSCP	HSP   HBFP
9	LFPR00009	ACTN4	O43707	EXOC	VSCP	HSP   HBFP
10	LFPR00010	ADGRG2	Q8IZP9	EXOC	VSCP	HSP   HBFP
11	LFPR00011	AGRN	O00468	EXOC	VSCP	HSP   HBFP
12	LFPR00012	AHNAK	Q09666	EXOC	VSCP	HSP   HBFP
13	LFPR00013	ANPEP	P15144	EXOC	VSCP	HSP   HBFP
14	LFPR00014	ALB	P02768	EXOC	VSCP	HSP   HBFP

Figura 15. 15 primeres files de la base de dades de proteïnes en la saliva de Life Fluids.

Posteriorment de l'ordenació de proteïnes, trobem un total de 286 proteïnes que s'han trobat en les 4 bases de dades d'on s'han extret proteïnes (apartat 3.3) en saliva. Destacar les 5 primeres proteïnes (amylase, cystatin S, mucin 7, prolactin -induced protein i zymogen granule protein) que també es van trobar en la literatura de Lacombe et al.[30].

M'agradaria destacar que la Cystatin S s'investiga tant pel seu potencial paper en la patologia de la síndrome de Sjögren[80] primari com per la possible utilització com a biomarcador de la malaltia[81].

Resaltar que la combinació de 3 proteïnes que apareixen en totes les bases de dades de recerca (cystatin B, triose-phosphate isomerase i una proteïna anomenada "deleted in malignant tumor 1 protein") van aconseguir un 85% de sensibilitat i un 80% d'especificitat per a la detecció de càncer gàstric[82].

Respecte les **proteïnes de l'alè**, la base de dades conté un total de **355 proteïnes**. Tenint en compte que la HBFP (base de dades de referència) conté un total de 351, Trobem un increment del 1,14%.

En la base de dades ja ordenada, ens mostra un total de 145 proteïnes trobades a la vegada en la nostra "Gold Standard" i a la única font literària útil, Lacombe et al.[30].

En l'estudi Lacombe et al.[30], destaca que l'alfa-1-antitripsina podria ser un candidat a biomarcador a l'emfisema pulmonar, particularment en aquells casos on l'emfisema està associat amb una deficiència d'alfa-1-antitripsina[83].

Realçar també que en l'estudi López-Sanchez et al.[84], destaca dos proteïnes que estan a la nostra base de dades (Hornerin i Hemoglobina) com potencial biomarcadors pel càncer de pulmó. L'estudi Newton et al.[85] suggereixen que l'hemoglobina pot tenir un paper en el metabolisme de l'òxid nítric i la protecció contra l'estrès oxidatiu i el nitrosatiu, amb implicacions importants en la fisiologia i la patologia pulmonar.

### 8.4 Elements químics

Pel que fa als **elements químics**, encara que el nombre d'elements identificats a la saliva no va augmentar significativament respecte la HMDB, el procés de filtratge i validació de la informació va ser fonamental per garantir la precisió en la creació de la nova base de dades. Les bases de dades d'elements químics tan de la saliva (Figura 16) com de l'alè, ha sigut ordenades en funció del seu nombre a la taula periòdica.

	A	B	C	D	E	F	G	H	I	J	K	L	
1	NumberID	Nom	Symbol	ecular We	inchikey	HMDB_ID	ubChem_I	KEGG_ID	pf_Element	le_of_Elem	Database	Literature	
2	LFCE003	Lithium	Li	6,941	HBBGRAR'	HMDB000		28486		Metal	Alkali Met	HMDB	LTMB0001
3	LFCE005	Boron	B	10,811	FRWDHM'	HMDB000	5462311			Metalloid	Metalloids	HMDB	
4	LFCE009	Fluoride	F	18,9984	KRHYFGT	HMDB000		28179	C00742	Non-met	Halogens	HMDB	LTMB0001
5	LFCE011	Sodium	Na	22,9898	FKNQFGJ	HMDB000		923	C01330	Metal	Alkali Met	HMDB	LTMB0001
6	LFCE012	Magnesium	Mg	24,305	JLVVSXFLK	HMDB000		888	C00305	Metal	Alkaline E:	HMDB	LTMB0001
7	LFCE013	Aluminum	Al	26,9815	XAGFODP'	HMDB000	5359268			Metal	Other Met	HMDB	
8	LFCE015	Phosphorus	P	30,9738	OAICVXFJF	HMDB000	5462309			Non-met	Non-met	HMDB	LTMB0001
9	LFCE017	Chlorine	Cl	35,453	VEXZGXHN	HMDB000		312	C00698	Non-met	Halogens	HMDB	LTMB0001
10	LFCE019	Potassium	K	39,0983	NPYPAHLE	HMDB000		813	C00238	Metal	Alkali Met	HMDB	

Figura 16. 10 primeres files de la base de dades d'elements químics en saliva de Life Fluids.

Subratllar que, dins d'aquest conjunt, solament un element, el Ferro, es va poder identificar en una altra base de dades diferent de la HMDB, en aquest cas MarkerDB[37]. Del total de elements trobats, 23 d'ells han aparegut com a mínim en una font literària.

Pel que fa a l'alè, hem trobat un total de 3 elements (Hidrògen, Oxigen i Argó), esmentats en l'aparat 6. Destacar la troballa d'Argó, un gas noble, en un estudi de Metabolomic Workbench (LFMB0084) [46]. S'ha investigat l'ús de l'Argó en el tractament de diverses patologies, incloent-hi la isquèmia cerebral, la lesió cerebral traumàtica i l'encefalopatia hipòxic-isquèmica. Tot i això, els models aplicats i l'administració del gas terapèutic varien entre els estudis[86].

En total s'han trobat 43 elements químics (Figura 17), i cap d'ells en ambdós biofluids.

Figura 17. Elements químics trobats en la Saliva (contorn de l'element en blau fosc), i l'Alè (contorn de l'element en blau clar: H, O, Ar).



## 8.5 Correlacions Saliva-Alè

Aquesta correlació pot revelar la complexitat de com interactua el cos humà amb el seu entorn i respon a les influències externes, especialment en el context de l'exposició a la contaminació del trànsit.

La saliva i l'alè són fonts d'informació valuoses sobre l'exposició de l'individu a les toxines ambientals i la resposta biològica. Tots dos es poden recollir de manera no invasiva, cosa que els converteix en eines atractives per als estudis epidemiològics i de salut[75].

En la correlació entre els dos biofluids en **metabòlits** ens ha identificat un total de 518 en comú com hem pogut veure en l'apartat 7.1. Tenim un total de **3.342 compostos identificats entre els dos biofluids** (1.877 de la saliva i 1.465 de l'alè), però només 518 que corresponen al **16%** són metabòlits que apareixen en les dos taules. Un 27,6% pel que fa a la taula de metabòlits a la saliva i un 36,6% respecte al total de metabòlits que trobem a la taula de l'alè.

En les taules **proteïnes** tenim un total de **5.165 compostos identificats entre els dos biofluids** (4.810 de la saliva i 355 de l'alè), però només 321 que corresponen al **6,2%**, són proteïnes que apareixen en les dos taules. Un 6,7% pel que fa a la taula de proteïnes a la saliva i un 90,4% respecte al total de metabòlits que trobem a la taula de l'alè. Resultat esperable si es té en compte la diferència de proteïnes de cada biofluid.

Estudiar les **vies metabòliques** és crucial per entendre la correlació entre els compostos de la saliva i l'alè. Això és perquè aquests compostos són productes dels processos bioquímics del cos, i la seva anàlisi pot revelar informació sobre els processos fisiològics que es duen a terme a l'organisme. Comprendre les vies metabòliques ajuda a identificar aquests biomarcadors i facilita el diagnòstic. Finalment, ja que la saliva i l'alè contenen productes de les vies metabòliques de la microbiota oral i pulmonar, l'anàlisi d'aquestes vies (Figura 18) pot aportar informació valuosa sobre la composició i funció de la microbiota, un factor clau per a la salut i la malaltia.

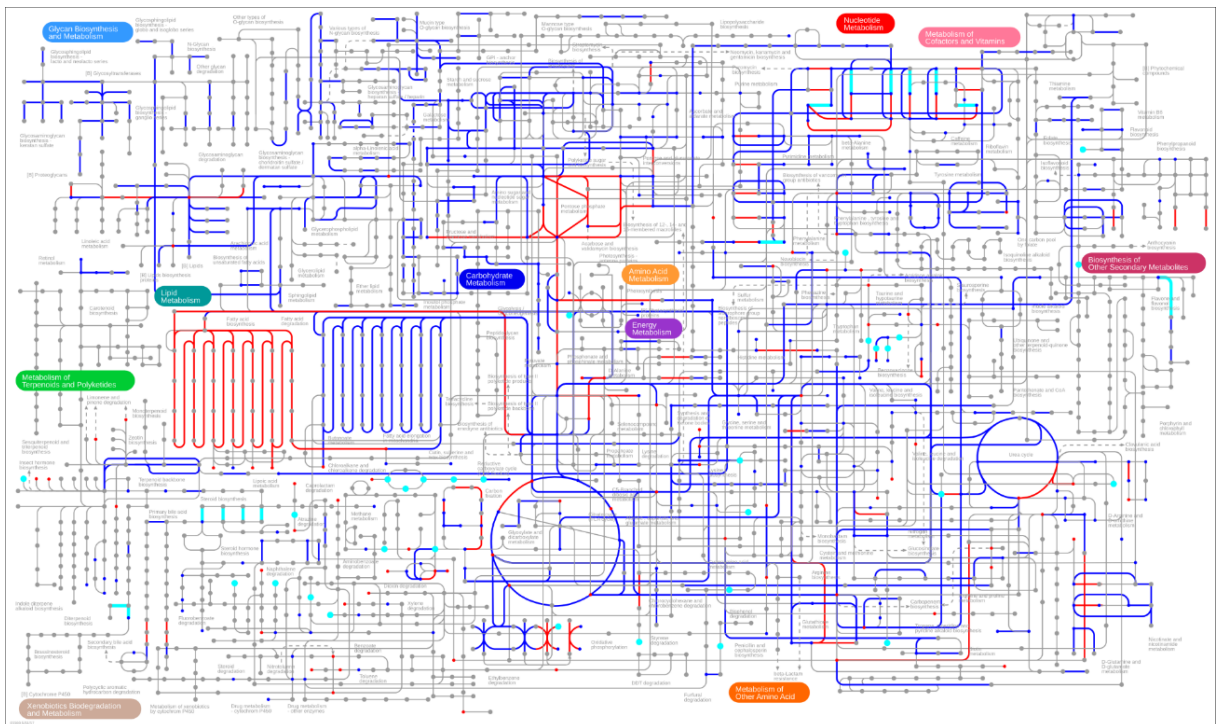


Figura 18. Il·lustració de les vies metabòliques de la saliva (blau fosc), l'alè (blau cel), i aquelles conjuntes (vermell).

La Figura 18 ens mostra les vies metabòliques (*metabolic pathways*) de les nostres bases de dades (Life Fluids Saliva i Life Fluids Alè), color blau fosc per els compostos de la saliva, blau clar per els de l'alè i les vies marcades en vermell són les vies comunes amb les compostos que s'han aconseguit amb la correlació.

En el nostre anàlisis **destaquen tres vies metabòliques** fonamentals en comú: el metabolisme dels lípids amb especial atenció a la biosíntesi d'àcids grassos, el metabolisme dels carbohidrats centrat en el metabolisme de la pentosa fosfat, i el metabolisme dels nucleòtids. El metabolisme dels lípids, i més específicament la biosíntesi d'àcids grassos, és un procés essencial a les cèl·lules humanes. Els àcids grassos són vitals en la formació de triglicèrids i fosfolípids, components essencials de les membranes cel·lulars, i són una font important d'energia.

A més, hem identificat la importància del metabolisme de la pentosa fosfat, una via paral·lela a la glucòlisi, en la composició química de la saliva i l'alè. Aquesta via metabòlica produeix ribosa-5-fosfat per a la síntesi de nucleòtids i NADPH, un potent reductor implicat en diverses reaccions bioquímiques, incloent-hi la síntesi de lípids. L'estudi Stincone et al. 2015[87] ha associat alteracions en aquesta via a malalties infeccioses o en la proliferació de les cèl·lules cancerígenes.

Finalment, el metabolisme dels nucleòtids, que exerceix un paper crucial en la vida de les cèl·lules com a components fonamentals de l'ADN i de l'ARN, també ha estat identificat al nostre estudi. L'alteració del metabolisme dels nucleòtids impulsa l'evasió immune de les cèl·lules canceroses [88].

## 8.6 Limitacions

Tot i aquests avenços, hem trobat una sèrie de limitacions durant la realització d'aquest estudi que han influït en els nostres resultats i que poden oferir orientació per a futurs esforços de recerca en aquest camp. La manca de literatura i estudis específics sobre els compostos presents sobretot en l'alè ha estat una limitació considerable, ja que ha limitat el nombre de compostos que hem pogut incloure a la nostra base de dades. En comparació amb biofluids més estudiats com la sang i l'orina, hi ha una manca evident d'informació sobre l'alè, cosa que subratlla la necessitat d'una investigació més gran en aquesta àrea.

## 8.7 Disponibilitat i protecció de les dades generades

Les bases de dades creades durant aquest TFG de la saliva i l'alè compten amb 8.550 entrades en total, repartides segons s'especifica a la Taula 7 (veure conclusions).

Les taules que componen les dues bases de dades Life Fluids de Saliva i de l'Alè, es troben **disponibles** pels membres del tribunal al drive de la URV:

[LifeFluidsSalivaAle\\_2023Juny](#)<sup>1</sup>

Les taules no s'adjunten en la memòria donat el volum total dels fitxers.

La base de dades Life Fluids per Saliva i Alè s'ha **registrat** a Safe Creative a data 12 de Juny de 2023, amb referència: 2306124568135.

---

<sup>1</sup> [https://rovira-my.sharepoint.com/:f:/g/personal/47790036-s\\_epp\\_urv\\_cat/Eq\\_HGt0pUUVOIL-GwGbXh1QBn7b7k-AtX4kzip8-1p3BCQ?e=GhwSPw](https://rovira-my.sharepoint.com/:f:/g/personal/47790036-s_epp_urv_cat/Eq_HGt0pUUVOIL-GwGbXh1QBn7b7k-AtX4kzip8-1p3BCQ?e=GhwSPw)

## 9 Conclusions

En aquest treball de fi de grau, s'ha assolit l'objectiu principal de desenvolupar una base de dades unificada que incorpora compostos de la saliva i l'alè, dos biofluids humans claus. Hem aconseguit identificar i unificar diverses fonts de dades, organitzar la nova base de dades "Life Fluids" i identificar components comuns entre la saliva i l'alè.

Hem constatat l'efectivitat de la nostra metodologia per superar les discrepàncies de nomenclatura, emprant identificadors específics, com ara InChIKey per a metabòlits i elements químics, i Uniprot per a proteïnes. Aquesta estratègia ha estat complementada amb l'addició d'identificadors públics àmpliament utilitzats, que proporcionen informació útil als investigadors.

Resultats: La nova base de dades "Life Fluids" inclou 1.877 metabòlits, 4.810 proteïnes i 40 elements químics en saliva; i 1.465 metabòlits, 355 proteïnes i 3 elements químics a l'alè. Tanmateix, hem de reconèixer que hi ha limitacions en el nostre treball, ja que la quantitat de compostos detectats dependrà de la qualitat i l'exhaustivitat de les fonts de dades originals.

Taula 7. Resum de les taules Life Fluids Saliva i Alè creades.

Òmica	Saliva (# entrades)	Alè (# entrades)	Total
Metabòlits	1.877	1.465	3.342
Proteïnes + gens associats	4.810	355	5.165
Elements químics	40	3	43
Total	6.727	1.823	8.550

La base de dades unificada "Life Fluids" representa una contribució significativa a la recerca biomèdica. Proveirà als investigadors una eina valuosa que els permetrà fer anàlisis més profunds i comparatius dels components de la saliva i l'alè. Aquesta informació pot ajudar a millorar la nostra comprensió dels processos biològics normals i patològics que es reflecteixen en aquests biofluids.

El treball presentat en aquest document marca un pas important cap a la consolidació de la informació existent sobre els components de la saliva i l'alè humà, i la metodologia emprada podria ser aplicada en altres estudis similars. Com a futura línia de recerca, suggerim l'expansió d'aquesta base de dades per incloure més compostos i per considerar altres biofluids.

## 10 Annex

A) Increment de metabòlits a la saliva.

Base de dades	Nombre d'entrades	Percentatge augmentat
HMDB	1.195	
Saliva_LFMB	1.877	57%

$$\text{Càlcul del increment} = \frac{LF - DB}{DB} = \frac{1.877 - 1.195}{1.195} = 0,57 = 57\%$$

B) Increment de metabòlits a l'alè.

Base de dades	Nombre d'entrades	Percentatge augmentat
HMDB	338	
Saliva_LFMB	1.465	333%

$$\text{Càlcul del increment} = \frac{LF - DB}{DB} = \frac{1.465 - 338}{338} = 3,33 = 333\%$$

C) Increment de proteïnes a la saliva.

Base de dades	Nombre d'entrades	Percentatge augmentat
HSP	3.442	
Saliva_LFMB	4.810	40%

$$\text{Càlcul del increment} = \frac{LF - DB}{DB} = \frac{4.810 - 3.442}{3.442} = 0,40 = 40\%$$

D) Increment de proteïnes a l'alè.

Base de dades	Nombre d'entrades	Percentatge augmentat
HBFP	351	
Saliva_LFMB	355	1,14%

$$\text{Càlcul del increment} = \frac{LF - DB}{DB} = \frac{355 - 351}{351} = 0,0114 = 1,14\%$$

## 11 Referències

- [1] Z. Huang *et al.*, "Saliva - A new opportunity for fluid biopsy," *Clinical Chemistry and Laboratory Medicine*, vol. 61, no. 1. De Gruyter Open Ltd, pp. 4–32, Jan. 01, 2023. doi: 10.1515/cclm-2022-0793.
- [2] A. Sharma, R. Kumar, and P. Varadwaj, "Smelling the Disease: Diagnostic Potential of Breath Analysis," *Molecular Diagnosis and Therapy*. Adis, May 01, 2023. doi: 10.1007/s40291-023-00640-7.
- [3] "Indicaciones, métodos diagnósticos y tratamiento erradicador de *Helicobacter pylori*: Recomendaciones de la II Conferencia Española de Consenso." [https://scielo.isciii.es/scielo.php?script=sci\\_arttext&pid=S1130-01082005000500007](https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1130-01082005000500007) (accessed May 30, 2023).
- [4] R. J. Keogh and J. C. Riches, "The Use of Breath Analysis in the Management of Lung Cancer: Is It Ready for Primetime?," *Current Oncology*, vol. 29, no. 10, pp. 7355–7378, Sep. 2022, doi: 10.3390/currenol29100578.
- [5] K. Dixit, S. Fardindoost, A. Ravishankara, N. Tasnim, and M. Hoorfar, "Exhaled Breath Analysis for Diabetes Diagnosis and Monitoring: Relevance, Challenges and Possibilities," *Biosensors (Basel)*, vol. 11, no. 12, p. 476, Nov. 2021, doi: 10.3390/bios11120476.
- [6] R. P. Horgan and L. C. Kenny, "'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics," *The Obstetrician & Gynaecologist*, vol. 13, no. 3, pp. 189–195, Jul. 2011, doi: 10.1576/toag.13.3.189.27672.
- [7] M. D. Adams *et al.*, "Complementary DNA Sequencing: Expressed Sequence Tags and Human Genome Project," *Science (1979)*, vol. 252, no. 5013, pp. 1651–1656, Jun. 1991, doi: 10.1126/science.2047873.
- [8] E. S. Lander *et al.*, "Initial sequencing and analysis of the human genome," *Nature*, vol. 409, no. 6822, pp. 860–921, Feb. 2001, doi: 10.1038/35057062.
- [9] J. Shendure and H. Ji, "Next-generation DNA sequencing," *Nat Biotechnol*, vol. 26, no. 10, pp. 1135–1145, Oct. 2008, doi: 10.1038/nbt1486.
- [10] Z. Wang, M. Gerstein, and M. Snyder, "RNA-Seq: a revolutionary tool for transcriptomics," *Nat Rev Genet*, vol. 10, no. 1, pp. 57–63, Jan. 2009, doi: 10.1038/nrg2484.
- [11] B. Domon and R. Aebersold, "Mass Spectrometry and Protein Analysis," *Science (1979)*, vol. 312, no. 5771, pp. 212–217, Apr. 2006, doi: 10.1126/science.1124619.
- [12] A. I. Nesvizhskii, "Proteogenomics: concepts, applications and computational strategies," *Nat Methods*, vol. 11, no. 11, pp. 1114–1125, Nov. 2014, doi: 10.1038/nmeth.3144.
- [13] D. S. Wishart, "Emerging applications of metabolomics in drug discovery and precision medicine," *Nat Rev Drug Discov*, vol. 15, no. 7, pp. 473–484, Jul. 2016, doi: 10.1038/nrd.2016.32.
- [14] "Systems Biology as Defined by NIH | NIH Intramural Research Program." <https://irp.nih.gov/catalyst/19/6/systems-biology-as-defined-by-nih> (accessed May 24, 2023).
- [15] "Biofluids." <https://lifefluids.com/es/biofluids/> (accessed May 25, 2023).

- [16] "Human Metabolome Database: About the Human Metabolome Database." <https://hmdb.ca/about> (accessed May 25, 2023).
- [17] "ExoCarta: Home - Exosome database." <http://exocarta.org/index.html> (accessed May 25, 2023).
- [18] "KEGG: Kyoto Encyclopedia of Genes and Genomes." <https://www.genome.jp/kegg/> (accessed May 25, 2023).
- [19] "BinVestigate - The BinBase investigation tool." <https://binvestigate.fiehnlab.ucdavis.edu/#/> (accessed May 26, 2023).
- [20] "MarkerDB." <https://markerdb.ca/about> (accessed May 26, 2023).
- [21] N. Drabińska *et al.*, "A literature survey of all volatiles from healthy human breath and bodily fluids: The human volatilome," *Journal of Breath Research*, vol. 15, no. 3. IOP Publishing Ltd, Apr. 01, 2021. doi: 10.1088/1752-7163/abf1d0.
- [22] J. A. M. Pereira, P. Porto-Figueira, R. Taware, P. Sukul, S. Rapole, and J. S. Câmara, "Unravelling the Potential of Salivary Volatile Metabolites in Oral Diseases. A Review," *Molecules*, vol. 25, no. 13. MDPI AG, Jul. 01, 2020. doi: 10.3390/molecules25133098.
- [23] A. Gardner, G. Carpenter, and P. W. So, "Salivary metabolomics: From diagnostic biomarker discovery to investigating biological function," *Metabolites*, vol. 10, no. 2. MDPI AG, Feb. 01, 2020. doi: 10.3390/metabo10020047.
- [24] Z. Huang *et al.*, "Saliva - A new opportunity for fluid biopsy," *Clinical Chemistry and Laboratory Medicine*, vol. 61, no. 1. De Gruyter Open Ltd, pp. 4–32, Jan. 01, 2023. doi: 10.1515/cclm-2022-0793.
- [25] "Metabolomics Workbench : Home." <https://www.metabolomicsworkbench.org/> (accessed May 26, 2023).
- [26] "MetaboLights - Metabolomics experiments and derived information." <https://www.ebi.ac.uk/metabolights/index> (accessed May 26, 2023).
- [27] "Vesiclepedia: Home - Extracellular vesicles database." <http://www.microvesicles.org/> (accessed May 26, 2023).
- [28] "HBFP." <https://bmbi.bmi.osumc.edu/HBFP/index.php/Home/Index/download.html?ftype%5B%5D=2> (accessed May 26, 2023).
- [29] "Category:Salivary Proteins - Human Salivary Proteome Wiki." [https://www.salivaryproteome.org/public/index.php/Category:Salivary\\_Proteins](https://www.salivaryproteome.org/public/index.php/Category:Salivary_Proteins) (accessed May 26, 2023).
- [30] M. Lacombe *et al.*, "Proteomic characterization of human exhaled breath condensate," *J Breath Res*, vol. 12, no. 2, Apr. 2018, doi: 10.1088/1752-7163/aa9e71.
- [31] "KNAPSAcK Core System." [http://www.knapsackfamily.com/knapsack\\_core/top.php](http://www.knapsackfamily.com/knapsack_core/top.php) (accessed May 26, 2023).
- [32] A. Sharma, R. Kumar, and P. Varadwaj, "Smelling the Disease: Diagnostic Potential of Breath Analysis," *Molecular Diagnosis and Therapy*. Adis, May 01, 2023. doi: 10.1007/s40291-023-00640-7.
- [33] D. S. Wishart *et al.*, "HMDB 5.0: the Human Metabolome Database for 2022," *Nucleic Acids Res*, vol. 50, no. D1, pp. D622–D631, Jan. 2022, doi: 10.1093/nar/gkab1062.
- [34] D. S. Wishart *et al.*, "HMDB: the Human Metabolome Database," *Nucleic Acids Res*, vol. 35, no. Database, pp. D521–D526, Jan. 2007, doi: 10.1093/nar/gkl923.

- [35] O. Fiehn, G. Wohlgemuth, and M. Scholz, "Setup and Annotation of Metabolomic Experiments by Integrating Biological and Mass Spectrometric Metadata," 2005, pp. 224–239. doi: 10.1007/11530084\_18.
- [36] Z. Lai *et al.*, "Identifying metabolites by integrating metabolome databases with mass spectrometry cheminformatics," *Nat Methods*, vol. 15, no. 1, pp. 53–56, Jan. 2018, doi: 10.1038/nmeth.4512.
- [37] D. S. Wishart *et al.*, "MarkerDB: an online database of molecular biomarkers," *Nucleic Acids Res*, vol. 49, no. D1, pp. D1259–D1267, Jan. 2021, doi: 10.1093/nar/gkaa1067.
- [38] F. M. Afendi *et al.*, "KNAPSAck Family Databases: Integrated Metabolite–Plant Species Databases for Multifaceted Plant Research," *Plant Cell Physiol*, vol. 53, no. 2, pp. e1–e1, Feb. 2012, doi: 10.1093/pcp/pcr165.
- [39] "Metabolomics Workbench : NIH Data Repository."  
<https://www.metabolomicsworkbench.org/data/DRCCMetadata.php?Mode=Study&StudyID=ST000418&StudyType=MS&ResultType=1> (accessed May 31, 2023).
- [40] "Metabolomics Workbench : NIH Data Repository."  
<https://www.metabolomicsworkbench.org/data/DRCCMetadata.php?Mode=Study&StudyID=ST000496> (accessed May 31, 2023).
- [41] "Metabolomics Workbench : NIH Data Repository."  
<https://www.metabolomicsworkbench.org/data/DRCCMetadata.php?Mode=Study&StudyID=ST000588> (accessed May 31, 2023).
- [42] "Metabolomics Workbench : NIH Data Repository."  
<https://www.metabolomicsworkbench.org/data/DRCCMetadata.php?Mode=Study&StudyID=ST001905> (accessed May 31, 2023).
- [43] "Metabolomics Workbench : NIH Data Repository."  
<https://www.metabolomicsworkbench.org/data/DRCCMetadata.php?Mode=Study&StudyID=ST002265> (accessed May 31, 2023).
- [44] "Metabolomics Workbench : NIH Data Repository."  
<https://www.metabolomicsworkbench.org/data/DRCCMetadata.php?Mode=Study&StudyID=ST002328> (accessed May 31, 2023).
- [45] "Metabolomics Workbench : NIH Data Repository."  
<https://www.metabolomicsworkbench.org/data/DRCCMetadata.php?Mode=Study&StudyID=ST000587> (accessed May 31, 2023).
- [46] "Metabolomics Workbench : NIH Data Repository."  
<https://www.metabolomicsworkbench.org/data/DRCCMetadata.php?Mode=Study&StudyID=ST000883> (accessed May 31, 2023).
- [47] N. S. Kale *et al.*, "MetaboLights: An Open-Access Database Repository for Metabolomics Data," *Curr Protoc Bioinformatics*, vol. 53, no. 1, Mar. 2016, doi: 10.1002/0471250953.bi1413s53.
- [48] "MTBLS2967 | 'Fuel for the Damage Induced': Untargeted Metabolomics in Elite Rugby Union Match Play." <https://www.ebi.ac.uk/metabolights/editor/MTBLS2967/descriptors> (accessed May 31, 2023).
- [49] "MTBLS2108 | Human age-declined saliva metabolic markers determined by LC-MS." <https://www.ebi.ac.uk/metabolights/editor/MTBLS2108/descriptors> (accessed May 31, 2023).
- [50] "MTBLS760 | Volatilomic insight of head and neck cancer via the effects observed on saliva metabolites." <https://www.ebi.ac.uk/metabolights/editor/MTBLS760/descriptors> (accessed May 31, 2023).

- [51] "MTBLS656 | A Pilot Characterization of the Human Chronobiome." <https://www.ebi.ac.uk/metabolights/editor/MTBLS656/descriptors> (accessed May 31, 2023).
- [52] "MTBLS524 | Identification of a discriminative metabolomic fingerprint of potential clinical relevance in saliva of patients with periodontitis using 1H nuclear magnetic resonance (NMR) spectroscopy." <https://www.ebi.ac.uk/metabolights/editor/MTBLS524/descriptors> (accessed May 31, 2023).
- [53] "MTBLS100 | The Human Saliva Metabolome." <https://www.ebi.ac.uk/metabolights/editor/MTBLS100/descriptors> (accessed May 31, 2023).
- [54] "MTBLS6760 | UHPLC-MS/MS-Based Identity Confirmation of Amino Acids Involved in Response to and Side Effects from Antiseizure Medications." <https://www.ebi.ac.uk/metabolights/editor/MTBLS6760/descriptors> (accessed May 31, 2023).
- [55] "Welcome to MassIVE." <https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp> (accessed May 30, 2023).
- [56] "PRIDE - Proteomics Identification Database." <https://www.ebi.ac.uk/pride/> (accessed May 30, 2023).
- [57] S. Keerthikumar *et al.*, "ExoCarta: A Web-Based Compendium of Exosomal Cargo," *J Mol Biol*, vol. 428, no. 4, pp. 688–692, Feb. 2016, doi: 10.1016/j.jmb.2015.09.019.
- [58] M. Pathan *et al.*, "Vesiclepedia 2019: a compendium of RNA, proteins, lipids and metabolites in extracellular vesicles," *Nucleic Acids Res*, vol. 47, no. D1, pp. D516–D519, Jan. 2019, doi: 10.1093/nar/gky1029.
- [59] D. Shao *et al.*, "HBFP: a new repository for human body fluid proteome," *Database*, vol. 2021, Oct. 2021, doi: 10.1093/database/baab065.
- [60] W. W. Lau, M. Hardt, Y. H. Zhang, M. Freire, and S. Ruhl, "The Human Salivary Proteome Wiki: A Community-Driven Research Platform," *J Dent Res*, vol. 100, no. 13, pp. 1510–1519, Dec. 2021, doi: 10.1177/00220345211014432.
- [61] S. R. Heller, A. McNaught, I. Pletnev, S. Stein, and D. Tchekhovskoi, "InChI, the IUPAC International Chemical Identifier," *J Cheminform*, vol. 7, no. 1, p. 23, Dec. 2015, doi: 10.1186/s13321-015-0068-4.
- [62] "Molecular weight | chemistry | Britannica." <https://www.britannica.com/science/molecular-weight> (accessed May 30, 2023).
- [63] R. M. Hartshorn *et al.*, "Brief Guide to the Nomenclature of Inorganic Chemistry", Accessed: May 29, 2023. [Online]. Available: <http://www.degruyter.com/pac>;
- [64] S. R. Heller, A. McNaught, I. Pletnev, S. Stein, and D. Tchekhovskoi, "InChI, the IUPAC International Chemical Identifier," *J Cheminform*, vol. 7, no. 1, p. 23, Dec. 2015, doi: 10.1186/s13321-015-0068-4.
- [65] S. Kim *et al.*, "PubChem 2019 update: improved access to chemical data," *Nucleic Acids Res*, vol. 47, no. D1, pp. D1102–D1109, Jan. 2019, doi: 10.1093/nar/gky1033.
- [66] "About NIH | National Institutes of Health (NIH)." <https://www.nih.gov/about-nih> (accessed May 29, 2023).
- [67] Z. Lai *et al.*, "Identifying epimetabolites by integrating metabolome databases with mass spectrometry cheminformatics," *Nat Methods*, vol. 15, no. 1, p. 53, Jan. 2018, doi: 10.1038/NMETH.4512.



- [68] M. Sud *et al.*, "LMSD: LIPID MAPS structure database," *Nucleic Acids Res*, vol. 35, no. Database, pp. D527–D532, Jan. 2007, doi: 10.1093/nar/gkl838.
- [69] D. S. Wishart *et al.*, "DrugBank: a knowledgebase for drugs, drug actions and drug targets," *Nucleic Acids Res*, vol. 36, no. suppl\_1, pp. D901–D906, Jan. 2008, doi: 10.1093/nar/gkm958.
- [70] "How do genes direct the production of proteins?: MedlinePlus Genetics." <https://medlineplus.gov/genetics/understanding/howgeneswork/makingprotein/> (accessed May 29, 2023).
- [71] "CTS Proxy." <https://cts.fiehnlab.ucdavis.edu/batch> (accessed May 30, 2023).
- [72] "Batch Compound Classification." <https://cfb.fiehnlab.ucdavis.edu/> (accessed May 30, 2023).
- [73] "Periodic Table – Royal Society of Chemistry." <https://www.rsc.org/periodic-table> (accessed May 30, 2023).
- [74] P. Wei Koh and P. Liang, "Understanding Black-box Predictions via Influence Functions," 2017.
- [75] C. N. Ladva *et al.*, "Metabolomic profiles of plasma, exhaled breath condensate, and saliva are correlated with potential for air toxics detection," *J Breath Res*, vol. 12, no. 1, p. 016008, Dec. 2017, doi: 10.1088/1752-7163/aa863c.
- [76] "LIPID MAPS." <https://www.lipidmaps.org/databases> (accessed Jun. 08, 2023).
- [77] D. S. Wishart *et al.*, "DrugBank 5.0: a major update to the DrugBank database for 2018," *Nucleic Acids Res*, vol. 46, no. D1, pp. D1074–D1082, Jan. 2018, doi: 10.1093/nar/gkx1037.
- [78] J. Li *et al.*, "Investigation of potential breath biomarkers for the early diagnosis of breast cancer using gas chromatography–mass spectrometry," *Clinica Chimica Acta*, vol. 436, pp. 59–67, Sep. 2014, doi: 10.1016/J.CCA.2014.04.030.
- [79] M. Mandy, F. Cornelia, L. Malgorzata, S. Oliver, S. Achim, and S. Dorothee, "P R A C E O R Y G I N A L N E Volatile organic compounds (VOCs) in exhaled breath of patients with breast cancer in a clinical setting Lotne związki organiczne (VOCs) w wydychanym powietrzu pacjentek z rakiem piersi w warunkach klinicznych," *Ginekol Pol*, vol. 83, pp. 730–736, 2012.
- [80] G. Cafaro *et al.*, "One year in review 2021: Sjögren's syndrome," *Clin Exp Rheumatol*, vol. 39, no. 6, pp. 3–13, Dec. 2021, doi: 10.55563/clinexprheumatol/eojaol.
- [81] D. Martini *et al.*, "Cystatin S—a candidate biomarker for severity of submandibular gland involvement in Sjögren's syndrome," *Rheumatology*, vol. 56, no. 6, pp. 1031–1038, Jun. 2017, doi: 10.1093/rheumatology/kew501.
- [82] M. Castagnola *et al.*, "Salivary biomarkers and proteomics: future diagnostic and clinical utilities," *Acta Otorhinolaryngologica Italica*, vol. 37, no. 2, pp. 94–101, Apr. 2017, doi: 10.14639/0392-100X-1598.
- [83] M. Fumagalli *et al.*, "Proteomic analysis of exhaled breath condensate from single patients with pulmonary emphysema associated to  $\alpha$ 1-antitrypsin deficiency," *J Proteomics*, vol. 71, no. 2, pp. 211–221, Jul. 2008, doi: 10.1016/j.jprot.2008.03.002.
- [84] L. M. López-Sánchez *et al.*, "Exhaled breath condensate biomarkers for the early diagnosis of lung cancer using proteomics," *American Journal of Physiology-Lung Cellular and Molecular Physiology*, vol. 313, no. 4, pp. L664–L676, Oct. 2017, doi: 10.1152/ajplung.00119.2017.

- [85] D. A. Newton, K. M. K. Rao, R. A. Dluhy, and J. E. Baatz, "Hemoglobin is expressed by alveolar epithelial cells," *Journal of Biological Chemistry*, vol. 281, no. 9, pp. 5668–5676, Mar. 2006, doi: 10.1074/jbc.M509314200.
- [86] A. Höllig, A. Schug, A. Fahlenkamp, R. Rossaint, M. Coburn, and Argon Organoprotective Network (AON), "Argon: Systematic Review on Neuro- and Organoprotective Properties of an 'Inert' Gas," *Int J Mol Sci*, vol. 15, no. 10, pp. 18175–18196, Oct. 2014, doi: 10.3390/ijms151018175.
- [87] A. Stincone *et al.*, "The return of metabolism: Biochemistry and physiology of the pentose phosphate pathway," *Biological Reviews*, vol. 90, no. 3, pp. 927–963, Aug. 2015, doi: 10.1111/BRV.12140.
- [88] N. J. Mullen and P. K. Singh, "Nucleotide metabolism: a pan-cancer metabolic dependency," *Nature Reviews Cancer* /, vol. 23, pp. 275–294, 2023, doi: 10.1038/s41568-023-00557-7.