

Treball de Fi de Grau

*Anàlisi del metiloma i el transcriptoma de
la diabetis de tipus 2*

Ariadna Cilleros Portet

Grau en Biotecnologia

Tutora: M.L.Calle

Vic, Juny del 2019

Vull agrair a M.L.Calle per haver-me guiat durant tot el procés del projecte. També agraeixo la col·laboració de l'Alba Casellas Comallonga, professora associada a l'UVIC i investigadora postdoctoral en el Centre d'Investigació Biomèdica en Red per la Diabetis i Malalties metabòliques Associades (CIBERDEM), i a la Lara Nonell, professora associada a l'UVIC i responsable de l'anàlisi de dades del Servei d'Anàlisi de Microarrays de l'Institut Hospital del Mar d'Investigacions Mèdiques.

ÍNDEX

RESUM	3
1.INTRODUCCIÓ	5
2.OBJECTIUS	6
3.CONTEXT	7
3.1.LA DIABETIS DE TIPUS 2	7
3.1.1.GENÈTICA DE LA DIABETIS DE TIPUS 2	8
3.2.EPIGENÈTICA I EXPRESSIÓ GÈNICA	9
4.METODOLOGIA	10
4.1.DADES	11
4.2.ANÀLISI DEL TRANSCRIPTOMA	12
4.3.ANÀLISI DEL METILOMA	14
5.RESULTATS I DISCUSSIÓ	16
5.1.TRANSCRIPTOMA	16
Insulin Like Growth Factor Binding Protein 1 (IGFB1)	19
ARN 5S RIBOSOMAL PSEUDOGEN 454	20
Sondes 7894124 i 7895099	22
5.2.METILOMA	24
Formyl Peptide Receptor 2	27
EH Domain Containing 4	28
Sterile Alpha Motif Domain Containing 11	28
SPARC/Osteonectin, Cwcv And Kazal Like Domains Proteoglycan 1	28
Keratin Associated Protein 24-1	29
Zinc Finger And BTB Domain Containing 22	29
ARFGEF Family Member 3	30
Tissue Inhibitor of Metalloproteinases 4 i Synapsin II	30
Nuclear Receptor Binding SET Domain Protein 1	31
HID1 Domain Containing	31
5.3.INTEGRACIÓ DELS DOS TIPUS DE DADES	32
6.CONCLUSIONS	34
7.PRESPECTIVES	35
8.BIBLIOGRAFIA	36
ANNEX	43

RESUM DEL TREBALL DE FINAL DE GRAU EN BIOTECNOLOGIA

Títol: *Anàlisi del metiloma i el transcriptoma de la diabetis de tipus 2*

Paraules clau: *diabetis de tipus 2 (DMT2), insulina, glucosa, metiloma, transcriptoma*

Autora: *Ariadna Cilleros Portet*

Tutora: *M.Luz Calle*

Data: *Maig del 2019*

La diabetis de tipus 2 (DMT2) o diabetes mellitus no insulíndependent és una malaltia metabòlica molt complexa que es caracteritza pel fet que les cèl·lules del nostre cos són incapaces de detectar la insulina a través dels seus receptors. El fet de desenvolupar DMT2 és el resultat de la interacció entre factors ambientals i factors heretables.

En aquest *Treball de Fi de Grau* hem volgut explorar alguns mecanismes moleculars que poden estar associats a la diabetis tipus 2 a partir d'una aproximació multi-òmica que integri informació epigenòmica i transcriptòmica. Per això, hem partit de les dades públiques de l'estudi (GSE65058): *Kirchner H, Sinha I, Gao H, Ruby MA et al. Altered DNA methylation of glycolytic and lipogenic genes in liver from obese and type 2 diabetic patients. Mol Metab 2016 Mar;5(3):171-183. PMID: 26977391*, que té per objectiu descobrir les contribucions epigenètiques en aquesta malaltia. Concretament, hem pogut analitzar els nivells de metilació i d'expressió gènica de pacients amb diabetis de Tipus 2 (DMT2) i controls sans "ND" obtingudes en aquest estudi.

Així doncs, l'objectiu de partida del present *Treball de Fi de Grau* és poder aportar llum sobre els mecanismes moleculars associats al desenvolupament de la diabetis de tipus 2 a partir de l'estudi de la possible relació entre el perfil d'expressió gènica i els nivells de metilació dels diabètics de tipus 2. Tant l'anàlisi del transcriptoma com el del metiloma, s'han dut a terme mitjançant el programa estadístic R, en el qual s'han utilitzat diverses llibreries de Bioconductor específiques per a cada tipus de dades.

En els resultats de l'anàlisi del transcriptoma vam identificar 4 sondes diferencialment expressades, on només una d'elles havia estat descrita anteriorment. En canvi pel metiloma diversos dinucleòtids CpG es trobaven diferencialment metilats. En el procés d'integració dels dos nivells d'informació molecular no vam poder identificar relacions rellevants, com per exemple dinucleòtids CpG que poguessin afectar l'expressió dels gens diferencialment expressats que havíem descrit en l'anàlisi del transcriptoma. De manera que aquests resultats deixen la porta oberta a noves vies d'investigació sobre les causes i les vies biològiques associades a la diabetis de tipus 2.

**SUMMARY DEGREE FINAL PROJECT
DEGREE IN BIOTECHNOLOGY**

Title: *Analysis of the methylome and transcriptome of the diabetes of type 2*

Keywords: *diabetes of type 2 (DMT2), insulin, glucose, methylome, transcriptome*

Author: *Ariadna Cilleros Portet*

Tutor: *M.Luz Calle*

Date: *May 2019*

The type 2 diabetes (DMT2) or diabetes mellitus non-insulin-dependent is a complex metabolic illness caused by the incompetence of the cells for detecting the insulin through its receptors. The development of the DMT2 is the result of the interaction between environmental factors and heritable factors.

In this *Degree Final Project*, we explored some molecular mechanisms that may be associated with the type 2 diabetes thanks to an approximation multi-omic which integrates the epigenetic and transcriptomic information. We started from the available public data of the study (GSE65058): *Kirchner H, Sinha I, Gao H, Ruby MA et al. Altered DNA methylation of glycolytic and lipogenic genes in liver from obese and type 2 diabetic patients. Mol Metab 2016 Mar;5(3):171-183. PMID: 26977391*, which objective is to discover the epigenetic contributions in the DMT2. In particular, we analyzed the methylation and gene expression levels of diabetics of type 2 (DMT2) and healthy controls “ND” used for Kirchner H. (2016).

The objective of this *Degree Final Project* is to shed light on the molecular mechanisms which are associated with the development of the type 2 diabetes doing an analysis of a possible relation between the methylation and gene expression levels of diabetics of type 2. Both analysis of the transcriptome and the methylome they had been carried out by the statistic program R, in which we used different specific libraries of the Bioconductor for each type of data.

On one hand, the results of the analysis of the transcriptome, we identified 4 differentially expressed probes, which only one of them had been described previously. On the other hand, analyzing the methylome we found many differentially methylated CpG sites. In the integration step, we couldn't identify relevant connections, such as CpG sites that could affect the expression of the differentially expressed genes that we described in the transcriptome analysis. Therefore, these results keep the door open to new research lines about the causes and the biological ways associated with type 2 diabetes.

1. INTRODUCCIÓ

La diabetis és una malaltia que es coneix des de fa més de 2000 anys, però no va ser fins al 1935 que es van diferenciar els dos tipus de diabetis que es coneixen actualment; la diabetis de tipus I o diabetis insulíndependent, i diabetis de tipus II o diabetis no insulíndependent. Totes dues comparteixen el fet que els nivells de glucosa en sang són més elevats del que caldria esperar, però la diabetis de tipus I és una malaltia autoimmune que dóna lloc a la pèrdua total de la producció d'insulina per part de les cèl·lules β pancreàtiques [1]. En canvi, la diabetis de tipus II es dóna quan les cèl·lules del cos són incapaces de reconèixer la insulina que circula pel torrent sanguini a través dels seus receptors [2].

A causa de la falta de coneixement sobre les causes que provoquen l'aparició de la diabetis de tipus 2 i el meu interès per l'exploració i l'anàlisi de les dades òmiques, aquest projecte té com a objectiu principal poder aportar llum als mecanismes genètics i epigenètics que donen lloc a la diabetis de tipus 2 a través d'una anàlisi multi-òmica.

En aquest *Treball de Fi de Grau* es van utilitzar les dades públiques de transcriptoma i metiloma de diabètics de tipus 2 i de persones sanes analitzades a l'estudi Kirchner H et al.(2016). Els avenços en la tecnologia han donat lloc a una gran quantitat de dades òmiques. La integració d'aquestes dades ens ofereix l'oportunitat d'avaluar les interaccions moleculars en múltiples nivells funcionals i proporcionar una comprensió més completa de les vies biològiques implicades en diferents malalties [3].

2. OBJECTIUS

A causa de les nombroses incògnites que rodegen la base genètica de la diabetis de tipus II i el meu interès per l'anàlisi de dades òmiques, l'objectiu principal d'aquest projecte és poder aportar llum a les causes que provoquen o estan associades al desenvolupament de la diabetis de tipus 2. Així doncs els objectius d'aquest treball són:

1. Identificar els gens diferencialment expressats entre persones amb diabetis de tipus II i persones sanes.
2. Estudiar els gens diferencialment expressats.

3. Identificar els dinucleòtids CpGs diferencialment metilats entre l'ADN de les persones que patien diabetis de tipus II i el de persones sanes.
4. Estudiar els gens vinculats als punts CpGs diferencialment metilats de l'ADN.
5. Integrar els resultats de l'anàlisi del transcriptoma i del metiloma.

3. CONTEXT

3.1.LA DIABETIS DE TIPUS 2

La diabetis de tipus 2 (DMT2) o diabetes mellitus no insulíndependent és una malaltia metabòlica molt complexa i, segons la Fundació Espanyola per la Diabetis, actualment el 13.8% d'espanyols majors de 18 anys han set diagnosticats, cosa que equival a uns 5.3 milions de persones [4]. Però el 43% del total desconeixien que patien aquesta malaltia.

Com ja hem dit a la Introducció, la diabetis de tipus 2 es caracteritza pel fet que les cèl·lules del nostre cos són incapaces de detectar la insulina a través dels seus receptors. El fet que no la detectin fa que la glucosa no sigui absorbida i conseqüentment que aquesta es vagi acumulant al torrent sanguini. L'augment de glucosa en sang fa que les cèl·lules β pancreàtiques siguin sobreexplotades per tal de segregar més insulina, causant un període d'hiperinsulinèmia. Aquesta sobreexplotació provoca que, de forma gradual, aquestes cèl·lules vagin perdent la capacitat de produir insulina i els nivells d'aquesta acaben disminuint i provoquen una situació d'hiperglucèmia [2].

La majoria de persones se'ls detecta que pateixen diabetis de tipus 2 quan realitzen anàlisis rutinàries ja que sol provocar símptomes que poden passar desapercebuts. Alguns d'aquest símptomes serien moltes ganes d'orinar (poliúria), molta set (polidípsia), molta gana (polifàgia), pèrdua de pes i altres símptomes menys freqüents com el cansament, rampes a les cames i somnolència [5].

El fet de desenvolupar DMT2 és el resultat de la interacció entre factors ambientals i un factor heretable. Els factors ambientals de risc, coneguts fins al moment, són l'obesitat, una vida sedentària, l'estrès i el fet d'haver nascut amb sobrepès o per sota del pes esperat. Aquests factors ambientals juguen un paper fonamental per presentar la malaltia, però no tenen el mateix impacte en tothom, això fa pensar que hi pugui haver un factor genètic

heretable. Però tot i això, sembla que no s'acaba d'aconseguir descriure amb exactitud la base genètica d'aquesta malaltia.

3.1.1. GENÈTICA DE LA DIABETIS DE TIPUS 2

Segons Ali O. (2013) [6] es creu que la diabetis de tipus 2 és un tipus de malaltia poligènica, o sigui que es desenvolupa degut a la interacció complexa entre múltiples gens i factors ambientals. Encara no s'ha pogut acabar de descriure la forma en què aquests gens interaccionen entre ells i els factors ambientals. A diferència de la diabetis de tipus 1, que el seu risc genètic es concentra majoritàriament en una regió en concret del genoma, el component genètic de DMT2 no es concentra en una regió, sinó que sembla ser el resultat de la interacció de múltiples gens dispersos.

Per tal de descobrir quins són els possibles gens que provoquen aquesta malaltia es van realitzar diferents estudis que es troben documentats a l'article d'Ali O. (2013). Es parla de tres tipus d'estudis realitzats: *linkage studies*, *candidate gen studies* i *genome wide association studies*.

Els *linkage studies*, que miren aquells gens que tenen tendència a heretar-se de forma conjunta a causa de la seva localització tan pròxima entre ells, van revelar dos gens, el calpain 10 (*CAPN10*) i el factor de transcripció 7-like 2 (*TCF7L2*).

Per altra banda, els *candidate gen studies* estudien gens dels que ja es sospita que poden estar relacionats amb la malaltia. Quan aquests estudis es van realitzar, es van centrar en gens involucrats en el metabolisme de la glucosa, la secreció d'insulina, la senyalització i el metabolisme dels lípids. Els estudis van revelar els següents gens: peroxisome proliferator-activated receptor gamma (*PPARG*), insulin receptor substrate 1 (*IRS1*) and *IRS-2*, potassium inwardly-rectifying channel, subfamily J, member 11 (*KCNJ11*), Wolfram syndrome 1 (wolframin) (*WFS1*), HNF1 homeobox A (*HNF1A*), HNF1 homeobox B (*HNF1B*) and *HNF4A*. El que més va sorprendre d'aquest estudi és que molts gens relacionats amb la secreció de la insulina no estaven lligats al desenvolupament de la diabetis de tipus II.

Per últim, els *genome wide association studies* consisteixen en genotipar milers de polimorfismes de nucleòtids simples (SNPs) que es troben en *linkage disequilibrium*.

Aquests estudis van provocar un augment del nombre de variants potencialment involucrades en la diabetis de tipus 2, fins a més de 60. El gen TCF7L2, que ja havia set identificat en els *linkage studies*, va ser el més significatiu de tots, però a més a més, també van sorgir: hematopoietically expressed homeobox (*HHEX*), Solute carrier family 30 member 8 (*SLC30A8*), cyclin-dependent kinase inhibitor 2A/B (*CDKN2A/B*), insulin-like growth factor 2 mRNA binding protein 2 (*IGF2BP2*), CDK5 regulatory subunit associated protein 1-like 1 (*CDKAL1*), high mobility group AT-hook 2 (*HMG A2*), potassium voltage gated channel, KQT like subfamily, member 1 (*KCNQ11*) i Notch 2-ADAM metallopeptidase domain 30 (*NOTCH2-ADAM30*).

3.2.EPIGENÈTICA I EXPRESSIÓ GÈNICA

Aquest *Treball de Fi de Grau* es basa en una anàlisi multi-òmica, o sigui en l'anàlisi de dues òmiques diferents. Òmica és un neologisme provinent de l'anglès que en biologia molecular s'utilitza com a sufix per referir-se a l'estudi de la totalitat o d'un conjunt, com gens, organismes d'un ecosistema, proteïnes, entre d'altres [7]. En aquest cas hem analitzat l'epigenòmica i la transcriptòmica que diferencien els diabètics de tipus 2 respecte a les persones sanes.

L'epigenoma es defineix com un conjunt de modificacions químiques que experimenta l'ADN quan determinats compostos químics i proteïnes s'hi uneixen i poden alterar accions com l'activació o desactivació de gens, i el control de la producció de proteïnes en cèl·lules específiques. Quan els compostos químics s'uneixen a l'ADN es diu que han "marcat" el genoma [8]. Els canvis epigenètics es diferencien dels genètics perquè aquestes marques no canvien la seqüència de l'ADN i són reversibles [9].

Les principals marques epigenètiques són la metilació de l'ADN i en les modificacions de les histones (metilació i acetilació). En aquest projecte ens hem centrat exclusivament en el conjunt de metilacions del genoma de les cèl·lules pancreàtiques, aquesta totalitat s'anomena metiloma [10].

Principalment, les metilacions de l'ADN es basen en la transferència d'un grup metil de la molècula S-adenosyl-L-methionine (SAM) a la citosina del dinucleòtid CpG [11]. Els dinucleòtids CpG o CpG sites són regions de l'ADN on una citosina és seguida per una

guanina en una seqüència lineal de bases en direcció 5'→3' [12]. Els grups metil units a citosines dels elements reconeguts per factors de transcripció inhibeixen aquest reconeixement, això dona lloc a la reducció de l'activitat transcripcional [9]. Parlem d'hypermetilació quan hi ha un increment de la metilació dels residus de les citosines en els dinucleòtids CpG a prop de regions reguladores de l'ADN [13]. En canvi parlarem d'hypometilació quan els nivells de metilació es veuen reduïts respecte el patró específic d'aquell teixit [14].

El patró de metilació d'aquests dinucleòtids CpG ve condicionat per l'entorn fetal i l'estil de vida de cada persona. Hi ha diversos estudis que afirmen que una mala dieta, l'obesitat, el sedentarisme o l'envelliment poden provocar canvis en aquest patró epigenètic, causant diabetis de tipus 2 [15].

La transcriptòmica és l'estudi del conjunt de transcrits d'ARN que són produïts pel genoma sota unes condicions específiques o un tipus cel·lular específic. La comparació del transcriptoma permet la identificació dels gens diferencialment expressats en diferents condicions, com per exemple en diabetics de tipus 2 [16]. Tot i això, el transcriptoma no només està compost per ARN missatger, també conté una àmplia varietat d'ARNs no codificants (ncRNA) [17].

Els ncRNA són molècules d'ARN que un cop transcrites no es tradueixen a proteïnes. Tenim varis exemples d'ARNs no codificants, alguns d'ells són força coneguts, com el ARN ribosòmic (rRNA) i el ARN de transferència (tRNA). En canvi n'hi ha d'altres que no són tan coneguts, com els micro ARN (miRNA) i els long non-coding ARN (lncRNA) [18].

Els micro ARN o miRNA es veuen involucrats en diverses funcions com el desenvolupament, la ploriferació, la diferenciació i la mort de les cèl·lules [9]. A través de la seva complementarietat amb un mRNA impedeix que aquest es tradueixi i això acaba donant lloc al silenciament del gen [18]. En canvi els lncRNA tenen com a funció principal ser un senyal molecular per regular la transcripció en resposta a diversos estímuls [19]. S'ha vist que el mal funcionament d'aquests dos tipus de ncRNA poden donar lloc a càncer, malalties respiratòries i malalties neurodegeneratives [20].

Tal com dèiem a la introducció, la integració del metiloma i el transcriptoma ens permetrà veure possibles relacions entre els nivells de metilació i l'expressió dels gens.

4. METODOLOGIA

4.1.DADES

Les dades de les quals vam partir en aquest *Treball de Fi de Grau* formaven part dels resultats de l'anàlisi del transcriptoma i el metiloma de pacients amb diabetis de Tipus 2 (DMT2) i controls sans "ND" extrets de l'estudi (GSE65058): *Kirchner H, Sinha I, Gao H, Ruby MA et al. Altered DNA methylation of glycolytic and lipogenic genes in liver from obese and type 2 diabetic patients. Mol Metab 2016 Mar;5(3):171-183. PMID: 26977391*, que te per objectiu descobrir les contribucions epigenètiques en la diabetis de tipus 2.

En total es van analitzar les mostres del teixit hepàtic de 28 subjectes classificats segons si patien o no diabetis de tipus 2 i segons si eren obesos, a 17 dels quals se'ls va analitzar tant el metiloma com el transcriptoma. Les dades de les 28 mostres estaven recollides en un total de 69 fitxers, 48 tenien el format .idat, característic de les plataformes BeadChip d'Illumina, i els 21 restants amb el format .CEL són característic de les plataformes d'Affymetrix. Les dades del transcriptoma i el metiloma es troben disponibles a la base de dades Gene Expression Omnibus (GEO) amb el número de sèrie GSE64998 i GSE65057, respectivament.

Per una banda, es va analitzar el transcriptoma de 7 diabètics obesos (T2D-obese), 8 obesos no diabètics (ND-obese) i 6 persones que no eren ni diabètiques ni obesos (non-obese), a través de la plataforma Affymetrix Human Gene 1.1 ST Array [21].

Per altra banda, el metiloma es va analitzar en les mostres de 9 diabètics obesos (T2D-obese), 8 obesos no diabètics (ND-obese) i 7 persones que no eren ni diabètiques ni obesos (non-obese) a través de la plataforma Illumina HumanMethylation450 BeadChip [22].

El protocol d'anàlisi de les dades que he dut a terme en aquest treball consisteix en primer lloc en un estudi de cada òmica per separat seguit de la integració dels resultats de les dues anàlisis. Concretament, de l'anàlisi de transcriptòmica s'obtenen els gens diferencialment expressats i de l'anàlisi de metilació s'obtenen les posicions CpG diferencialment metilades. La integració dels resultats va consistir en, d'una banda, l'estudi dels nivells de metilació dels gens diferencialment expressats i d'altra banda, els

nivells d'expressió dels gens identificats en l'anàlisi del metiloma. A la [Figura 1](#) hi podem veure un esquema de l'anàlisi que es va portar a terme pel metiloma i el transcriptoma. Veiem que comparteixen varies etapes en comú, com el control de qualitat, la normalització de les dades i l'anotació de les sondes. La resta d'etapes depenien del tipus de dades i de la plataforma amb la qual s'havien analitzat. Per tal de fer l'anàlisi reproducible, tot el codi desenvolupat en aquest TFG es troba disponible en lliure accés al GitHub a la següent adreça <https://github.com/ariadnacillers/TFG>

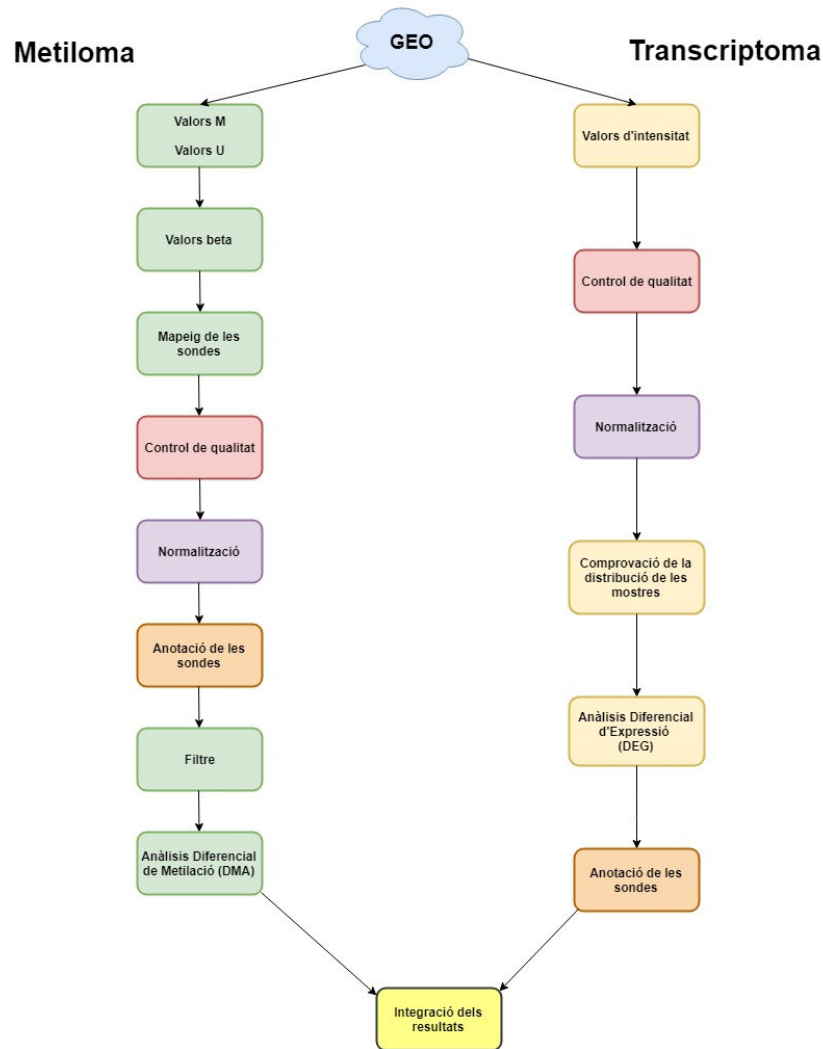


Figura 1: Diagrama de flux que es va seguir per realitzar l'anàlisi de les mostres del metiloma i del transcriptoma.

4.2. ANÀLISI DEL TRANSCRIPTOMA

Tal com hem dit a l'apartat anterior, pel transcriptoma es van analitzar un total de 21 mostres de teixit hepàtic: 7 diabètics obesos (T2D-obese), 8 obesos no diabètics (ND-obese) i 6 persones que no eren ni diabètiques ni obeses (non-obese). Aquestes mostres van ser analitzades amb la plataforma Human Gene 1.1 ST d'Affymetrix [21] que conté un total de 750.000 sondes úniques de 25 nucleòtids que interroguen més de 28.000 gens [23]. Aquesta plataforma es caracteritza per contenir sondes tant per mesurar ARN missatger (mRNA) com per mesurar transcrits d'ARN llargs no codificant (lncRNA) [24]. Els lncRNA són molècules d'ARN de més de 200 nucleòtids, que es troben involucrats en la remodelació de la cromatina i en la regulació de la transcripció i de la post-transcripció [25].

Com podem veure en la [Figura 1](#), a partir del valor de les intensitats de l'array el primer pas realitzat va ser el control de qualitat (QC) que ens serveix per comprovar la distribució de les dades i que no hi hagi cap mostra que s'hagi de descartar. En aquest pas es van visualitzar les imatges dels arrays i es van fer tres tipus de gràfics: un MAplot [26,27,28], un histograma i un boxplot.

Un MAplot és un gràfic de dispersió on en l'eix X hi trobem el logaritme de la mitjana d'intensitat de les mostres i en l'eix Y el logaritme de la distribució de les intensitats. Per tal de poder saber si una mostra es troba desviada o no, hem de fixar-nos en la línia vermella. Com menys curvatura mostri i més recta sigui la línia vermella, menys desviada estarà la mostra. Una forma de mirar-ho és a través dels valors de la mediana, com més propers siguin a 0, menys desviació pateix la mostra [29].

La normalització va ser el següent pas que es va portar a terme. És on es van corregir els biaixos sistemàtics de les dades, eliminar els efectes tècnics i agrupar els valors de les sondes en gens. Les dades del transcriptoma es van normalitzar a través de la funció *rma* [30] del paquet *oligo* de Bioconductor. Per comprovar els resultats es va tornar a fer l'histograma i el boxplot de la distribució dels nivells d'expressió.

Seguidament es va comprovar la distribució de les mostres, per tal de veure si s'observava algun tipus d'agrupació dels resultats, com per exemple, diabètics i no diabètics, a través del Principal Component Analysis de la llibreria *factoextra* [31].

Finalment, l'anàlisi d'expressió diferencial va consistir en un model prenent com a variable dependent l'expressió del gen i com a variable explicativa el fet de tenir o no diabetis de tipus 2, ajustat pel fet de patir o no obesitat, ja que l'obesitat és un factor de risc molt rellevant en aquesta malaltia. Aquest pas es va realitzar a través del model lineal del paquet de funcions de *limma* [32].

Per poder identificar els gens corresponents a les diferents sondes es va realitzar un pas d' anotació, on es van fer servir els paquets *annotate* [33] i *hugene10sttranscriptcluster.db* [34] de Bioconductor.

4.3. ANÀLISI DEL METILOMA

Tal com hem dit a l'apartat de les dades, pel metiloma es van analitzar un total de 24 mostres de teixit hepàtic: 9 diabètics obesos (T2D-obese), 8 obesos no diabètics (ND-obese) i 7 persones que no eren ni diabètiques ni obesos (non-obese). Aquestes mostres van ser analitzades amb la plataforma Illumina HumanMethylation450 BeadChip [21] que conté un total de 485.512 dinucleòtids CpG [35]. Illumina HumanMethylation450 BeadChip es basa en l'ús d'unes microesferes com a plataformes on s'uniran les seqüències d'ADN que prèviament hauran set tractades amb bisulfit, amplificades i fragmentades. El tractament amb bisulfit serveix per poder diferenciar les citosines metilades de les que no ho estan. Cada dinucleòtid CpG que s'interrogui tindrà dos tipus de boles: una per l'estat metilat de la citosina (M) i l'altre per l'estat desmetilat (U). A causa de l'amplificació que s'ha realitzat tenim diverses còpies de cada CpG, les còpies metilades s'uniran a M i les desmetilades a U. Seguidament es fa un cicle de síntesi d'un sol nucleòtid que ve marcat amb biotina, tant en les boles U com en les boles M del mateix locus CpG s'hi afegirà el mateix nucleòtid. Un cop fet el cicle es tenyirà l'array i s'escanejarà les intensitats de cada tipus de bola (U i M) [36]. D'aquesta manera obtindrem la intensitat de metilació i la intensitat de la no metilació.

Tal com podem observar en l'esquema del procés d'anàlisi del metiloma que trobem a la [Figura 1](#), a partir dels valors d'intensitats M i U (% metilació i % no metilació) per cada individu, obtenim els valors beta. Els valors beta, o valors de metilació de l'ADN, són

valors continus entre el 0 i l'1 que representen la proporció de metilació d'un locus CpG. Els valors beta estan definits com a:

$$Beta_i = \frac{\max(M_i, 0)}{\max(U_i, 0) + \max(M_i, 0) + \alpha}$$

On M_i i U_i són les intensitats mesurades per les sondes metilades i no metilades del dinucleòtid CpG i^{th} . Per evitar valors negatius després de la correcció del *background*, tots els valors negatius seran restablerts com a 0. Illumina recomana afegir una constant α (per defecte, $\alpha = 100$) al denominador per regularitzar el valor Beta quan les intensitats de les sondes metilades i no metilades són baixes [37].

Un cop vam tenir els valors beta, vam mapejar les sondes al genoma de referència a través de la funció `mapToGenome` del paquet *minfi* [38,39,40,41,42,43,44] de Bioconductor. El següent pas que es va portar a terme va ser un control de qualitat a través de la funció `runShinyMethyl` del paquet *ShinyMethyl* [45] de Bioconductor. Aquesta funció realitza un diagrama de les intensitats de les sondes tractades amb bisulfit, un gràfic representant la diferència de medianes de la intensitat de les metilacions i les no metilacions i una representació de les densitats dels valors beta.

Un cop realitzat el control de qualitat i comprovat que no s'ha de descartar cap mostra, es va passar al pas de la normalització a través de la funció `preprocessQuantile` del paquet *minfi* [38,39,40,41,42,43,44]. Aquesta funció implementa un procés de normalització per quantil per a microarrays de metilació d'Illumina en el que les sondes són estratificades per regió (locus CpG). Per comprovar els resultats es va tornar a realitzar un histograma de les densitats de les sondes metilades.

Tal com es va fer amb el transcriptoma, el model estadístic que es va utilitzar per detectar els locus CpG diferencialment metilat va ser el mateix: la variable dependent en aquest cas són els valors beta que indiquen el grau de metilació i la variable explicativa el fet de patir o no diabetis de tipus 2, ajustat pel fet de patir o no obesitat. Com ja s'ha dit aquest pas es va realitzar a través del model lineal del paquet de funcions de *limma* [32].

Finalment, per poder obtenir més informació de les sondes, aquestes es van anotar amb el paquet *IlluminaHumanMethylation450kanno.ilmn12.hg19* [46]. De tots els locus CpG obtinguts es van descartar els que pertanyien als cromosomes sexuals, per tal d'evitar desviacions dels resultats deguts al gènere, i aquelles regions que no es trobessin o en el promotor o en el cos del gen.

5. RESULTATS I DISCUSSIÓ

5.1. TRANSCRIPTOMA

Com veiem a la Figura A1 (annex) els arrays es mostren intactes sense desplaçament de les sondes ni contaminants. Respecte a la Figura A2 (annex) les línies vermelles de les diferents mostres són majoritàriament rectes i no mostren curvatures. A més a més, els valors de la mediana són pròxims a 0 de manera que les mostres no semblen patir una desviació.

La Figura 2 és un histograma que ens mostra la distribució dels nivells d'expressió de les dades no normalitzades i en canvi la Figura 3 ens mostra la distribució de les dades un cop normalitzades. Com podem veure si comparem la Figura 3 amb la Figura 2, les dades normalitzades es troben distribuïdes de manera més uniforme gràcies al fet que s'han corregit els biaixos sistemàtics de les dades.

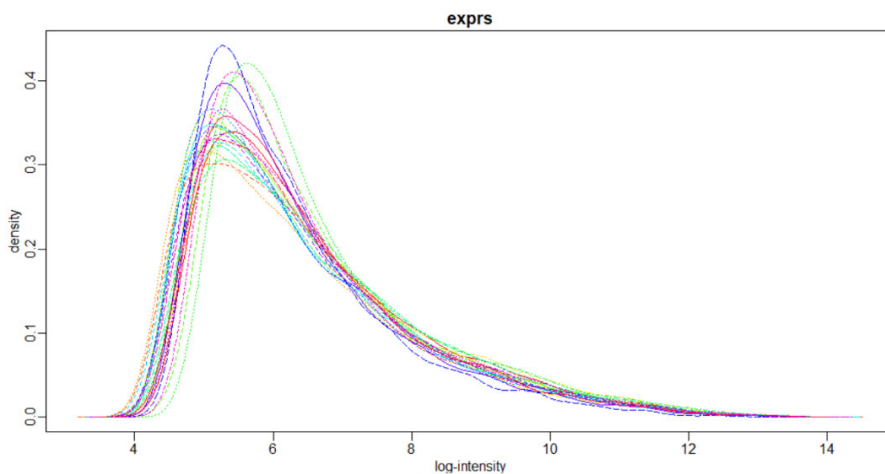


Figura 2: Histograma de la distribució dels nivells d'expressió de les dades no normalitzades del transcriptoma. En l'eix y hi trobem les densitats, mentre que a l'eix x el logaritme de les intensitats de les mostres.

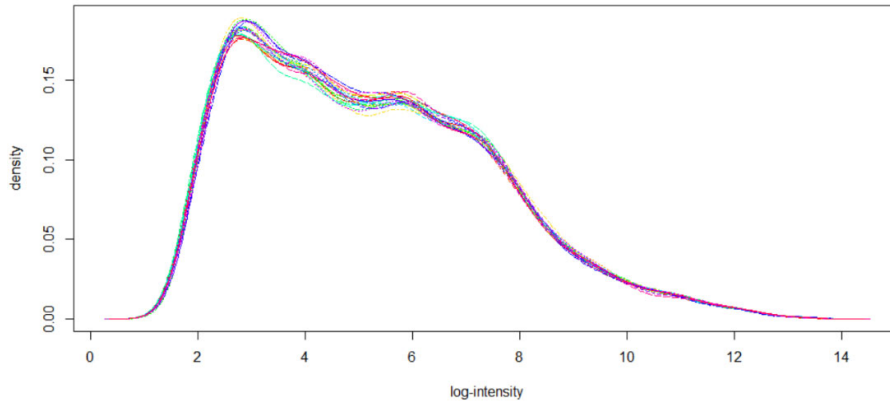


Figura 3: Histograma de la distribució dels nivells d'expressió de les dades normalitzades del transcriptoma. En l'eix y hi trobem les densitats, mentre que a l'eix x el logaritme de les intensitats de les mostres.

La Figura 4 és un boxplot de la distribució dels nivells d'expressió de les mostres no normalitzades i la Figura 5 ens mostra la mateixa distribució però amb les dades normalitzades. Com podem veure si comparem l'alçada de les caixes de les mostres normalitzades i de les no normalitzades, veiem que les caixes de les diverses mostres normalitzades es troben totes a una alçada quasi idèntica gràcies a la correcció dels biaixos sistemàtics de les dades.

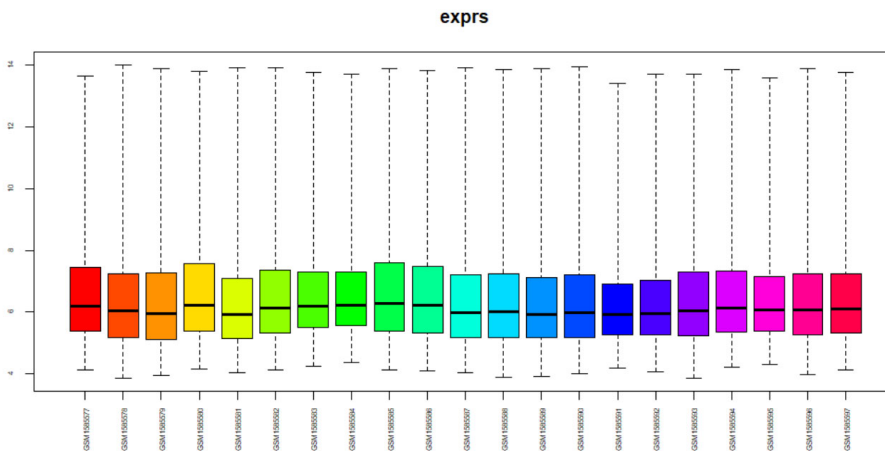


Figura 4: Boxplot de la distribució dels nivells d'expressió de les mostres no normalitzades del transcriptoma. En l'eix de les x hi trobem les mostres i en l'eix y hi trobem el logaritme de de les intensitats.

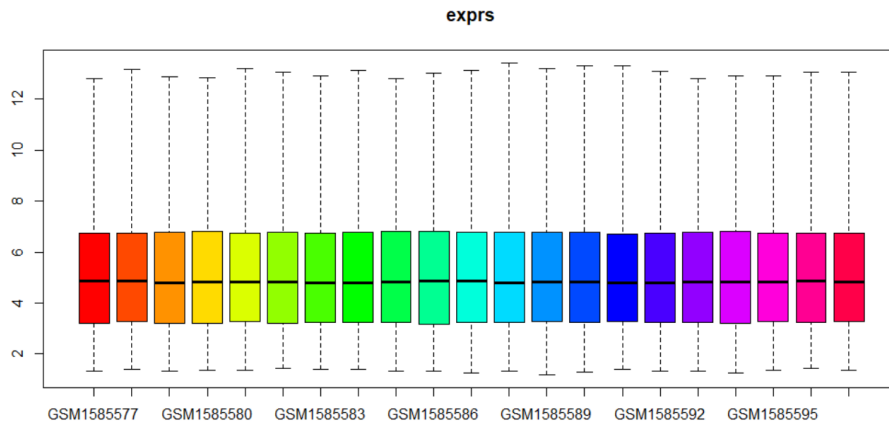


Figura 5: Boxplot de la distribució dels nivells d'expressió de les mostres normalitzades del transcriptoma. En l'eix de les x hi trobem les mostres i en l'eix y hi trobem el logaritme de les intensitats.

En la Figura 6, observem que la Dim2 separa en dos grups els no diabètics no obesos (ND-NO) a la part superior del gràfic i els no diabètics obesos (ND-O) i els obsessos diabètics (D-O) a la part inferior de l'eix. Respecte a la Dim1, la majoria dels ND-NO queden a l'esquerra d'aquesta component, mentre que la majoria dels D-O es troben a la dreta de l'eix. En canvi els ND-O es troben escampats al llarg dels dos quadrants.

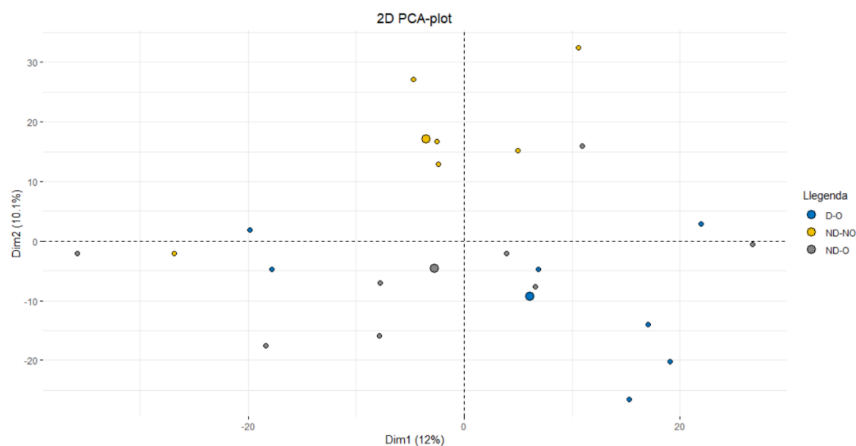


Figura 6: Principal Component Analysis de les mostres normalitzades del transcriptoma. La Dim1 representa la component principal 1 i la Dim2 és la component principal 2. La Dim1 explica el 12% de la variabilitat de les mostres, en canvi la Dim2 n'explica el 10.1%.

Amb l'anàlisi del transcriptoma vam identificar 4 sondes diferencialment expressades (Taula 1), considerant com a llindar un p-valor inferior a 0.05 i un valor absolut de log Fold Change (logFC) superior a 1. El logFC es defineix com el logaritme de la proporció

entre dos valors [47]. En el nostre cas aquests dos valors són el nivell d'expressió mitjà d'un gen en diabètics i en no diabètics i el logaritme es pren en base 2. D'aquesta manera el logFC ens indica la quantitat de vegades en què el gen s'expressa més en un grup que en l'altre. Per exemple, un logFC=1 indica que el gen s'expressa el doble (2^1) en diabètics que en no diabètics, un logFC=2, indica que el gen s'expressa 4 vegades (2^2) més en diabètics que en no diabètics, etc. En canvi, un logFC negatiu indica que el gen s'expressa més en el grup de no diabètics que en el de diabètics. Per exemple, un logFC= -1 indica que el gen s'expressa el doble en no diabètics que en diabètics, etc. Normalment es consideren gens diferencialment expressats aquells amb un logFC en valor absolut superior a 2. En el nostre cas no vam trobar cap sonda que complís aquesta condició i vam decidir reportar aquells resultats amb un logFC en valor absolut superior a 1.

De les 4 sondes identificades, només dues d'aquestes corresponen a gens, les dues primeres són sondes intròniques control de les que ja parlarem més endavant. Pel que fa als gens, per una banda tenim el pseudogen ribosomal 5S (RN5S454) i per altra banda, l'Insulin Like Growth Factor Binding Protein 1 (IGFBP1) que descriurem a continuació.

Taula 1: Sondes diferencialment expressades entre diabètics i no diabètics.

Probe ID	Símbol	Gene name	Chr	logFC	p-valor
7895099	NA	NA	NA	-1.80	0.00
7894124	NA	NA	NA	-1.60	0.00
8021043	RN5S454	Pseudogen ribosomal 5S	18	1.60	0.01
8132694	IGFBP1	Insulin Like Growth Factor Binding Protein 1	7	-1.80	0.02

Insulin Like Growth Factor Binding Protein 1 (IGFBP1)

L'IGF1 és produït en resposta de l'hormona del creixement circulant en sang. Les seves funcions principals són estimular el creixement del cos i promoure el creixement de les cèl·lules del múscul esquelètic, el cartílag, l'os, el fetge, els nervis, la pell, les cèl·lules

hematopoètiques i el pulmó [48]. A més a més de promoure la captació de la glucosa com la insulina, l'IGF1 també pot regular la síntesi de l'ADN cel·lular [49].

L'IGF1 pot circular per la sang de forma lliure o unit a IGFBP1. Quan els Insulin Like Growth Factor estan units a la IGFBP1 els impedeixen portar a terme la seva activitat [50].

En els nostres resultats, els valor de logFC ens indica que IGFBP1 es troba reprimat en el grup de diabètics. Els nivells baixos d'IGFBP1 estan associats amb la incidència de la diabetis de tipus 2 [51]. En un període d'hiperinsulinèmia, com el que es viu en la diabetis de tipus 2, la insulina tendeix a inhibir l'expressió d'IGFBP1 per evitar que aquesta s'uneixi amb IGF1. Provocant que hi hagi la IGF1 circulant lliure per la sang, fa que aquesta pugui portar a terme les seves funcions, entre elles l'estimulació de la captació de glucosa. Per aquest motiu les persones que pateixen diabetis de tipus 2, tenen un nivell inferior d'IGFBP1 [52].

Recentment s'ha descrit que la gent amb obesitat i/o diabètics de tipus 2 tenen un risc molt més elevat de patir diversos càncers. Una de les principals causes són les vies de senyalització que desencadena la unió de l'IGF1 amb el seu receptor cel·lular [53].

ARN 5S RIBOSOMAL PSEUDOGEN 454

La tercera sonda correspon al pseudogen ribosomal 454. Un pseudogen és un fragment d'ADN que conté tots els elements que hauria de contenir un gen però no són funcionals com a tal, ja que pateixen alguna alteració. Els pseudogens poden haver-se originats a través d'un transcrit d'un gen gràcies a l'acció d'una transcriptasa reversa, o bé a través de la duplicació d'un gen [12].

L'ARN 5S ribosomal pseudogen 454 conté un exó, que en teoria no és codificant, i té una llargada de 127bp. Segons la versió 38 del genoma humà (GRCh38.p12) es troba en el cromosoma 18 a la coordenada 42270589..42270715 [54]. Gràcies a la tècnica RNAseq, se sap que es troba en el cervell, el còrtex, el cerebel, les artèries, la medul·la, el colon i les tiroides. No se'n sap la funció i no sembla que s'hagi relacionat amb la diabetis ni la insulina en cap estudi.

Respecte als nostres resultats, sembla que hi ha una sobreexpressió del gen en el grup de diabètics. Com ja hem dit, de moment no s'ha descrit cap mena de relació amb la diabetis

de tipus 2 ni la insulina, però el que més destaca de tot és la localització del pseudogen. Com veiem en l'esquema (Figura 7) de la seva localització, just uns parells de bases abans i en la mateixa orientació, hi trobem el gen de PIK3C3 (phosphatidylinositol 3-kinase catalytic subunit type 3). PIK3C3 és una subunitat catalítica de la PI3K, un complex enzimàtic que és activat pel receptor de la insulina i el de IGF1 un cop els seus lligands s'hi uneixen. Aquest complex s'encarrega de la síntesi de glicogen, la inhibició de la gluconeogènesi i la síntesi de proteïnes [55].

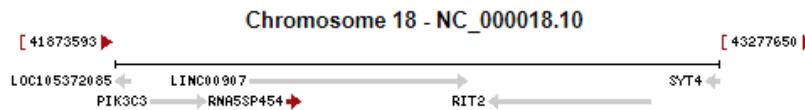


Figura 7: Esquema de la localització de RN5S454 [56]. Just a l'extrem 5' del pseudogen hi trobem el gen PIK3C3 que es llegeix en la mateixa direcció (5'→3') que RN5S454.

En concret el que succeeix quan la insulina s'uneix al seu receptor (Figura 8) és que aquest allibera un substrat (IRS1) que serà fosforilat i seguidament activarà el complex PI3K [57], que desencadenarà una cascada d'esdeveniments que acabaran resultant les funcions, que hem comentat abans.

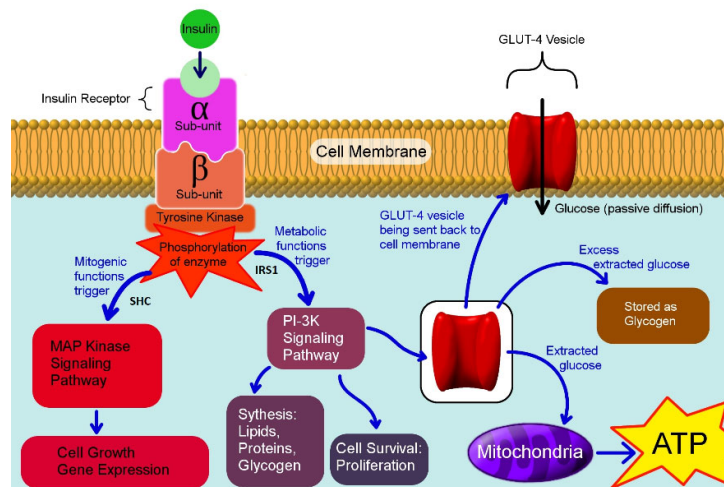


Figura 8: Esquema de la transducció de senyal del receptor de la insulina [57]. Podem veure que el substrat IRS1 del receptor de la insulina desencadena la via de senyalització PI3K, mentre que el substrat SHC desencadena la via de senyalització MAP quinasa.

Una possible explicació d'aquest fenomen, tot i que no tenim dades per poder-lo validar, podria ser que PIK3C3 estigués patint una mutació puntual de tipus readthrough. O sigui,

que en comptes de contenir el senyal de poliadenilació, aquesta hauria patit una mutació puntual de manera que la maquinària de la transcripció no la detectaria i la transcripció seguiria fins al següent senyal de poliadenilació, que correspondria al del pseudogen ribosomal 454. Això, si fos cert, provocaria que el transcrit de la subunitat PIK3C3 contingüés el transcrit del pseudogen i permetria explicar perquè els diabètics tenen una sobreexpressió d'aquest. A més a més, la interacció entre PI3K i IRS1, podria arribar a suposar un possible problema en la cadena de transducció de senyal del receptor de la insulina. Però aquesta teoria no acaba d'encaixar pel fet que la ruta PI3K està involucrada en molts processos de manera que tots aquests s'haurien de veure afectats d'alguna manera o altra.

Sondes 7894124 i 7895099

Les dues primeres sondes més diferencialment expressades són 7894124 i 7895099. En la plataforma Human Gene 1.1 ST d'Affymetrix cadascuna correspon a un *intronic normalization control (negative control)* [58,59]. Aquestes sondes contenen seqüències de regions intròniques de gens control que s'utilitzen posteriorment per normalitzar les dades [60,61,62], de manera que la seva intensitat d'expressió hauria de ser igual en totes les mostres.

Però en els nostres resultats sembla que dues d'aquestes sondes, la 7894124 i la 7895099, no s'estiguin expressant de la mateixa manera entre diabètics i els no diabètics. En concret són zones intròniques que es troben reprimides en el transcriptoma dels diabètics. A partir d'aquests punts es contemplen tres possibles explicacions: 1) que hi ha un gen que manté aquest intró en les persones sanes i en canvi els diabètics l'eliminen durant l'splicing, 2) que aquest intró codifiqui per un miRNA o lncRNA, o bé 3) que sigui degut a un error en el processament dels arrays.

Per tal de poder comprovar la primera hipòtesi hauríem de disposar de les coordenades de la sonda, però com que no ens la donen vam fer un blast amb les seqüències de les sondes. Com que són seqüències tan curtes i el nostre genoma és tan gran, en van sortir moltes coincidències i per tant no vam poder saber el gen en concret al qual pertanyia.

També hem dit que podrien codificar per un miRNA o un lncRNA. Per tal de comprovar si les nostres sondes formen part d'algun miRNA es va utilitzar la base de dades miRBase

[63,64,65,66,67,68]. La sonda 7895099 va ser alineada amb dos miRNA: hco-miR-5924-5p de l'espècie *Haemonchus contortus* i nle-miR-7180 de l'espècie *Nomascus leucogenys* (Taula 2). Els alineaments no coincideixen al 100%, sinó que hi havia dos mismatches per cadascun. L'expected value (Evalue) és un paràmetre que mesura la significació de l'alineament, com més s'aproxima a 0 més significatiu és l'alineament [69]. En el nostre cas els Evalue ens indiquen que l'alineament no és significatiu, per tant no podem confirmar que les nostres seqüències són ortòlegs d'aquests miRNA.

Taula 2: Alineament de la sonda 7895099.

Accession	ID	Strand	Score	Evalue	Alignment
MIMAT0023403	hco-miR-5924-5p	+	62	7.7	<pre> 8 cacagccagacgcggu 23 4 cacagccagaugcagu 19 </pre>
MIMAT0049422	nle-miR-7180	-	62	7.7	<pre> 5 guucacagccagacgc 20 18 guacacagccagaggc 3 </pre>

La sonda 7894124 té diverses seqüències però totes van ser alineades amb un mateix miRNA (Taula 3). Sembla que ens trobem una situació molt semblant a la de l'altre sonda, tenim un total de 3 mismatches i un score i una Evalue que ens indiquen que l'alineament no és significatiu. Com que aquest miRNA sí que pertany a l'*Homo sapiens*, en aquest cas no podríem confirmar que estiguessin relacionats. Per aquest miRNA no s'ha descrit que estigui vinculat a cap malaltia de moment [70].

Taula 3: Alineament de la sonda 7894124.

Accession	ID	Strand	Score	Evalue	Alignment
MIMAT0018190	hsa-miR-3916	+	63	6.3	<pre> 8 gaaacuucugguucucag 25 9 gaaauggcugguucucag 26 </pre>

L'altre possibilitat que s'havia contemplat és que les dues sondes corresponguessin a un ARN llarg no codificant. Per tal de poder saber-ho es va utilitzar la base de dades Incrnadb [71,72], però no va coincidir amb cap lncRNA descrit. De manera que en el cas que fos un ARN llarg no codificant encara no s'hauria descobert o bé no constaria en aquesta base de dades.

Per últim podria ser que aquests resultats fossin fruit d'un error en els arrays, en concret un *batch effect*. Els *batch effects* és el biaix tècnic que ha pogut produir-se durant la manipulació de les mostres. Per exemple, si has de marcar les mostres però en tens moltes com per fer-ho en el mateix moment, hauràs de fer-ho en diferents rondes de marcatge. Això provocaria que les mostres marcades al mateix moment tindran el mateix biaix tècnic, però aquelles mostres que han set marcades en diferents moments tindran un biaix diferent i per tant el pas de normalització no ens corregiria a la perfecció aquest biaix [73]. Com que no hem portat a terme el processament dels arrays i la preparació de les mostres, no podem descartar aquesta hipòtesi.

5.2. METILOMA

La Figura 9 és un diagrama de les diferències de les medianes de les intensitats metilades i no metilades de cada mostra no normalitzada. En aquest cas les mostres es troben força agrupades en clúster, de manera que no considerem que n'hi hagi cap que es desviï de forma significativa.

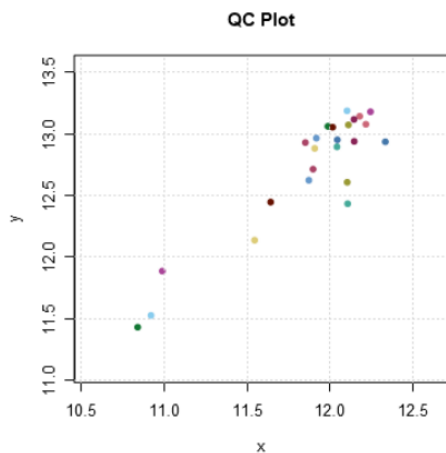


Figura 9: Diagrama del control de qualitat de la metilació. Cada punt es refereix a la diferència entre la mediana de les intensitats de metilació i no metilació de cada mostra.

La Figura 10 i la Figura 11 ens mostren les densitats dels valors beta de les dades no normalitzades i normalitzades, respectivament. Veiem que tant en una figura com en l'altra els valors beta són molt semblants entre les mostres i no sembla que hi hagi desviacions, però les dades normalitzades tenen unes alçades i unes amplitudes encara més semblants. Aquest fet es deu a la correcció dels efectes tècnics.

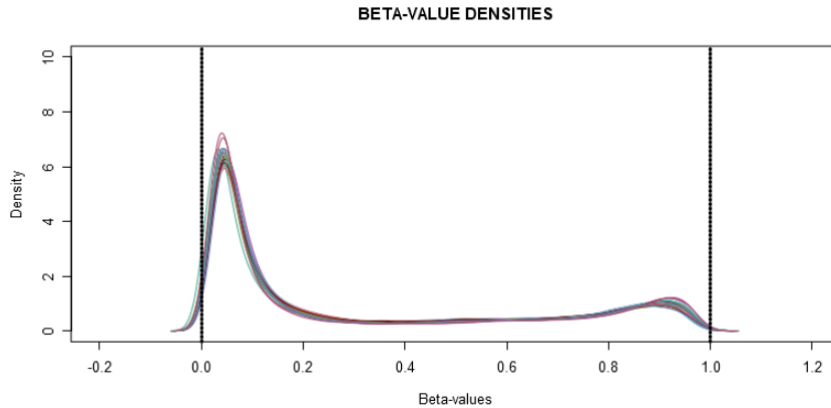


Figura 10: Densitats dels valors beta de les dades no normalitzades del metiloma. En l'eix y hi trobem les densitats, mentre que a l'eix x els valors beta.

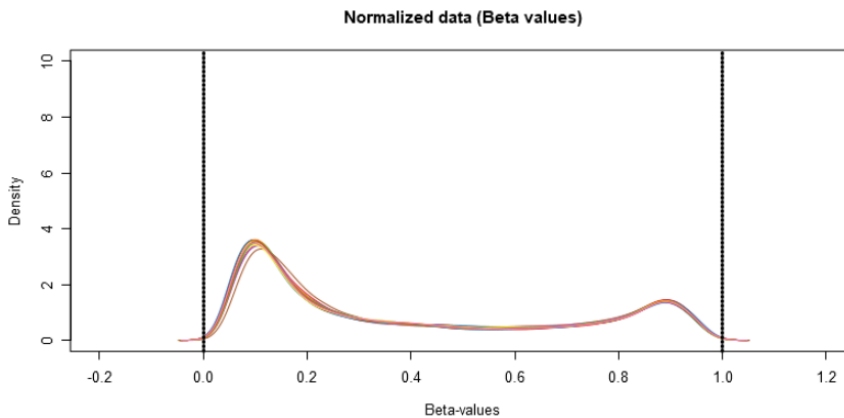


Figura 11: Densitats dels valors beta de les dades normalitzades del metiloma. En l'eix y hi trobem les densitats, mentre que a l'eix x els valors beta.

Amb l'anàlisi del metiloma vam identificar diversos punts CpGs diferencialment metilats, considerant com a llindar un p-valor inferior a 0.05. A continuació hem descrit els 10 primers dinucleòtids més significatius (amb el p-valor més petit) (Taula 4). A la Taula 4 les columnes indicades amb D i ND corresponen a la mitjana dels valor beta dels diabètics i dels no diabètics, respectivament. La columna indicada amb D-ND és la diferència entre

les dues mitjanes de valors betes, de manera que ens indica la diferència de la proporció de metilació entre els dos grups. Si aquest valor és positiu vol dir que el dinucleòtid CpG es troba més metilat en diabètics i si el valor és negatiu que es troba més metilat en no diabètics.

Taula 4: Resultats del metiloma.

CpG	Gene name	Chr	p-valor	D	ND	D-ND
cg25653204	FPR2	19	4.025816e-06	0.2146	0.3231	0.1085
cg09204108	EHD4	15	6.957319e-06	0.8577	0.8326	0.025
cg04121631	SAMD11	1	7.651091e-06	0.1802	0.2116	-0.0314
cg26968544	SPOCK1	5	1.063683e-05	0.6721	0.7399	-0.0678
cg24184737	KRTAP24-1	21	1.234854e-05	0.7632	0.8003	-0.0371
cg16544736	ZBTB22	6	1.364040e-05	0.2372	0.2920	-0.0548
cg12434085	KIAA1244	6	1.629013e-05	0.8753	0.8477	0.0276
cg10958002	TIMP4;SYN2	3	1.789415e-05	0.5332	0.5036	0.0296
cg23383189	NSD1	5	1.830179e-05	0.3149	0.2641	0.0508
cg02971219	C17orf28	17	1.883935e-05	0.3886	0.4274	-0.0388

Per tal de visualitzar els resultats del metiloma vam fer un volcano plot (Figura 12). Aquest tipus de gràfic permet visualitzar les sondes més diferencialment metilades tenint en compte el p-valor i la diferència de metilació. Cada punt representa un locus CpG i com més lluny del centre i més elevat es trobi més diferencialment metilada es trobarà aquella sonda.

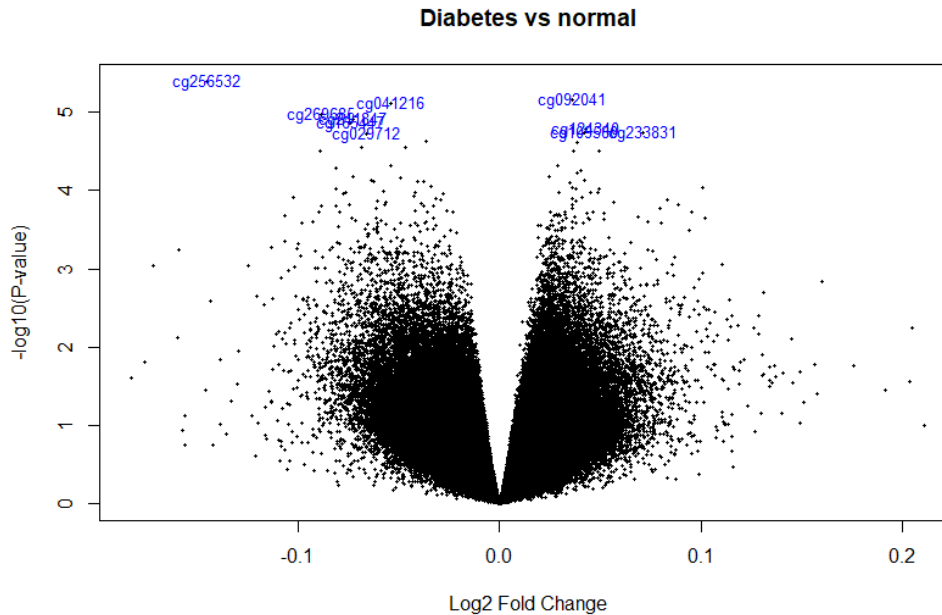


Figura 15: Volcano plot dels resultats del metiloma. En l'eix y hi trobem el logaritme del p-valor i en l'eix x la diferència de proporcions.

🚩 Formyl Peptide Receptor 2 (FPR2)

El dinucleòtid cg25653204 es troba en el gen Formyl Peptide Receptor 2 (FPR2), que es localitza al cromosoma 19 (51,752,026-51,770,670 de la versió GRCh38/hg38) i té un nombre total de 5 exons. Se sap que es pot unir a diversos lligands per això està relacionat a diverses vies de senyalització, com l'Akt, i en diversos sistemes, com el sistema immunitari innat. Està descrit que està relacionada a les malalties causades per prions [74] i l'any 2016 es va descobrir que podria ser un bon biomarcador per malalts amb tuberculosi pulmonària que patien diabetis de tipus 2 [75].

En concret la via de senyalització Akt és activada per la PI3K, de la que parlàvem en els resultats del transcriptoma, i és l'encarregada de promoure la supervivència de la cèl·lula i el creixement d'aquesta davant d'estímuls externs [76].

Pel que fa el nivell de metilació entre els dos grups veiem segons la Taula 4 que els no diabètics tenen un nivell de metilació més elevat respecte els diabètics.

EH Domain Containing 4

El locus cg09204108 es troba dins el gen EH Domain Containing 4 (EHD4), que es localitza en el cromosoma 15 (41,895,939-41,972,578 de la versió GRCh38/hg38) i té un nombre total de 6 exons. Es diu que probablement controla la reorganització de la membrana a través de la hidròlisi de l'ATP i el transport de l'endosoma en els primers estadis de vida. De moment no està relacionat amb cap desordre ni malaltia [77].

Respecte a la Taula 4 veiem que el nivell de metilació en diabètics és una mica més elevat, però els valors entre els dos grups són molt semblants.

Comparant aquest resultat amb els del transcriptoma, sabem que no es relaciona directament amb cap dels gens esmentats basant-nos en les vies de senyalització en les que EHD4 està implicat [78].

Sterile Alpha Motif Domain Containing 11

Aquest dinculeòtid CpG es troba en el gen Sterile Alpha Motif Domain Containing 11 (SAMD11), que es localitza en el cromosoma 1 (923,928-944,581 de la versió GRCh38/hg38) i té un nombre total de 14 exons. Es diu que desenvolupa un rol en el desenvolupament de fotoreceptors. Sembla que està relacionada amb la retinitis pigmentaria i amb la fundus dystrophy, però no amb la diabetis de tipus 2 [78].

A la Taula 4 veiem que aquest locus es troba més metilat en les persones no diabètiques.

De moment no s'ha descrit cap relació amb els gens esmentats en el transcriptoma, sembla ser que segons la base de dades GeneCards només interacciona amb la proteïna SAMD7 [78].

SPARC/Osteonectin, Cwcv And Kazal Like Domains Proteoglycan 1

El locus cg26968544 es troba en el gen SPARC/Osteonectin, Cwcv And Kazal Like Domains Proteoglycan 1 (SPOCK1), que es localitza en el cromosoma 5 (136,975,298-137,598,379 de la versió GRCh38/hg38) i conté 11 exons. Es creu que juga un rol important en les interaccions cèl·lula-cèl·lula i cèl·lula-matriu, i que contribueix a diversos mecanismes neuronals del sistema nerviós central. En la base de dades

GeneCards es diu que està relacionada amb la malaltia congenital chylothorax i hereditary coproporphyria, dues malalties orfes [79].

Respecte a la Taula 4, sembla que el nivell de metilació en diabètics és més baix que el de les persones sanes.

No sembla ser que es relacioni directament amb els gens descrits en el transcriptoma, però en aquest cas SPOCK1 interacciona amb IGF1 a través de la via de senyalització ERK, que es porta a terme per tal de comunicar els receptors extracel·lulars i el nucli cel·lular [79].

Keratin Associated Protein 24-1

Com s'indica a la Taula 4 aquest dinucleòtid CpG es troba en el gen Keratin Associated Protein 24-1 (KRTAP24-1), que es localitza en el cromosoma 21 (30,281,309-30,282,958 de la versió GRCh38/hg38) i conté 1 exó. Estan involucrats en la queratinització del cabell a través de realitzar enllaços disulfur. De moment no s'ha relacionat amb cap desordre [80].

Tenint en compte que els no diabètics són el grup control, els diabètics tenen un nivell inferior de metilació.

Segons l'entrada del gen a la base de dades GeneBank, sembla que fins al moment no s'ha descrit que la proteïna del gen KRTAP24-1 interaccioni amb els gens esmentats en el transcriptoma [80].

Zinc Finger And BTB Domain Containing 22

El dinucleòtid cg16544736 es troba en el gen Zinc Finger And BTB Domain Containing 22 (ZBTB22), que es localitza al cromosoma 6 (33,314,405-33,317,942 de la versió GRCh38/hg38) i conté 3 exons. Sembla que està involucrat en la regulació de la transcripció. A la base de dades GeneCards està documentat que el ZBTB22 està involucrat en diversos desordres; en la lobomiosis, l'anosogonia, penis squamous cell carcinoma, dermatographia i granulomatous amèbic encephalitis [81].

Tal com s'indica a la Taula 4, sembla que els diabètics tenen un nivell de metilació més baix que les persones sanes.

De moment no està documentat que estigui relacionat amb cap dels gens descrits en els resultats del transcriptoma [81].

ARFGEF Family Member 3

El dinucleòtid cg12434085 es troba en el gen ARFGEF Family Member 3 (KIAA1244), que es localitza en el cromosoma 6 (138,161,916-138,344,663 de la versió GRCh38/hg38) i conté fins a 34 exons. Participa en la regulació de l'homeòstasis de la glucosa, regulant de forma negativa la biogènesi dels grànuls d'insulina a les illes pancreàtiques, i regula la producció de glucagó a les cèl·lules α . Inhibeix la translocació de PHB2 (Repressor Of Estrogen Receptor Activity) i pot augmentar l'activitat transcripcional de l'estrogen alfa (ESR1) a les cèl·lules de càncer de mama. A més a més de la seva involucració amb el càncer de mama, no hi ha cap altre desordre que figuri en el seu perfil de les diverses bases de dades [82].

A la Taula 4 podem observar que els nivells de metilació en diabètics són més elevats que en no diabètics.

Actualment no està documentat que interactuï directament amb cap dels gens descrits en els resultats del transcriptoma, però està implicat en el metabolisme de proteïnes de la mateixa manera que IGFBP1 [83].

Tissue Inhibitor of Metalloproteinases 4 i Synapsin II

Sembla que aquest locus CpG es troba en dos gens; TIMP4 i SYN2. Per una banda, Tissue Inhibitor of Metalloproteinases 4 (TIMP4), es localitza en el cromosoma 3 (12,153,051-12,159,351 de la versió GRCh38/hg38) i conté 5 exons. És un inhibidor de la matriu de les metalloproteinases, un grup de peptidases involucrades en la degradació de la matriu extracel·lular. A la base de GeneCards està documentat que està relacionat amb el càncer de mama [84].

Per altra banda, synapsin II (SYN2) es troba en el cromosoma 3 (12,004,360-12,192,032 de la versió GRCh38/hg38), superposat amb el gen TIMP4, i conté 17 exons. Bàsicament

és una fosfoproteïna neuronal que cobreix les vesícules sinàptiques, s'uneix al citoesquelet i es creu que desenvolupa una funció en la regulació de l'alliberament de neurotransmissors. També pot exercir un paper en la secreció de la noradrenalina per les neurones del sistema nerviós simpàtic [85]. Està involucrat en l'esquizofrènia, el desordre bipolar i l'autisme.

Pel que fa el nivell de metilació, sembla que en diabètics és més elevat que en les persones que no pateixen diabetis de tipus 2.

Ni TIMP4 ni SYN2 semblen estar involucrades directament en cap via de senyalització en la que hi trobem les proteïnes dels gens del resultat del transcriptoma [84,85].

Nuclear Receptor Binding SET Domain Protein 1

El locus cg23383189 es troba en el gen Nuclear Receptor Binding SET Domain Protein 1 (NSD1), que es localitza en el cromosoma 5 (177,133,025-177,300,215 de la versió GRCh38/hg38) i conté fins a 30 exons. Bàsicament és una histona metiltransferasa, o sigui que transfereix grups metil a les histones. Depenen del context cel·lular, aquella metilació permetrà o no la transcripció. Està involucrat en la sotos syndrome 1, beckwith-wiedemann syndrome, preeclàmpsia/eclàmpsia 1, pre-eclàmpsia i l'hypertelorsim [86].

Pel que fa als resultats de la Taula 4, veiem que els diabètics tenen un nivell de metilació més elevat.

Segons la base de dades GeneCards no sembla que estigui descrit cap relació directa entre la proteïna del gen NSD1 i els gens resultants de l'anàlisi del transcriptoma [86].

HID1 Domain Containing

El dinucleòtid CpG cg02971219 es troba en el gen HID1 Domain Containing (C17orf28), que es localitza en el cromosoma 17 (74,950,742-74,973,166 de la versió GRCh38/hg38) i conté fins a 17 exons. Està documentat que aquest gen té un rol molt important en el desenvolupament de diversos càncers en diversos teixits [87].

Tal com indica la Taula 4, els diabètics presenten un nivell inferior de metilació en aquesta locus.

Segons la seva entrada en la base de dades GeneCard no s'ha descrit que s'hagi relacionat directament amb cap gen del resultat del transcriptoma [87].

En el cas que trobéssim diversos dinucleòtids CpG en una mateixa regió d'un gen podrien constituir una illa CpG. Si aquests dinucleòtids estiguessin diferencialment metilats dificultarien el pas de la maquinaria de transcripció i per tant, s'estaria reprimint l'expressió d'aquell gen. Però és la suma de tots aquests dinucleòtids que produiria l'efecte, per si sols no tenen suficient força [88]. De manera que, per tal de poder saber si els gens dels que hem parlat anteriorment es troben sobreexpressats o reprimits hauríem de mirar el nivell de metilació de diversos locus CpG.

5.3. INTEGRACIÓ DELS DOS TIPUS DE DADES

Per tal de poder aprofundir i integrar més els resultats anteriors s'han analitzat els nivells de metilació dels gens del resultat del transcriptoma i els nivells d'expressió dels gens del resultat del metiloma.

Per una banda, dels gens diferencialment expressats que hem identificat en l'anàlisi del transcriptoma, només hem pogut mirar l'estat de metilació dels locus CpGs que regulaven el gen IGFBP1 (Taula 5), ja que en aquest array del metiloma no s'analitzaven aquells locus CpG que es trobaven al pseudogen ribosomal 454.

Taula 5: Nivells de metilació dels locus CpGs que regulen el gen IGFBP1.

CpG	p-valor	D	ND	D-ND
cg01310829	0.374444	0.2163	0.2144	-0.0019
cg00110785	0.0589783	0.2759	0.2852	-0.0093
cg01650726	0.08065013	0.1843	0.1978	-0.0135
cg10326031	0.8279373	0.8765	0.8726	0.0039
cg10662959	0.0718485	0.1356	0.1475	-0.0119
cg11593949	0.2959011	0.5266	0.5366	-0.01
cg20237960	0.5378222	0.9123	0.9169	-0.0046

cg22108980	0.1408656	0.2097	0.2270	-0.0173
cg27447599	0.6268584	0.3414	0.3166	0.0248

Tal com veiem a la Taula 5, la majoria de locus CpGs tenen un nivell de metilació molt semblant entre els dos grups. Per això, considerant un p-valor de 0.05, les diferències del nivell d'expressió entre els dos grups no es consideren significatives. En el transcriptoma l'IGFB1 tenia un p-valor significatiu, en canvi en el metiloma els seus dinucleòtids CpG veiem que no es troben diferencialment metilats. El fet que en el transcriptoma es mostri una repressió de l'expressió i en el metiloma no es doni una metilació diferencial dels locus CpGs d'aquest gen, podria indicar que la seva expressió no està regulada a través de la metilació d'aquestes regions.

Per altra banda, també s'han analitzat els valors d'expressió dels gens que identificats en el metiloma (Taula 6), excepte el del gen NSD1 que no figurava en la llista de les sondes que havien pogut ser mapajades en l'anotació.

Taula 6: Nivells d'expressió dels gens dels resultats del metiloma.

Probe ID	Símbol	logFC	p-valor
8030860	FPR2	-0.10	0.27
7987772	EHD4	0.10	0.24
7896761	SAMD11	-0.10	0.48
8114287	SPOC1	0.00	0.71
8069800	KRTAP24-1	0.10	0.47
8178988	ZBTB22	0.00	0.53
8122279	KIAA1244	0.00	0.63
8085360	TIMP4	0.00	0.75
8077879	SYN2	0.10	0.37
8018264	C17orf28	-0.10	0.59

Tal com veiem a la Taula 6, els gens descrits en els resultats del metiloma tenen p-valors superiors al nivell de significació 0.05, de la mateixa manera que el metiloma, i els valors absoluts de logFC són inferiors a 1. Això indica que no són gens diferencialment expressats entre diabètics de tipus 2 i persones sanes. Per tant o bé no estan regulats a

través de la metilació dels seus locus CpG o bé necessiten varis CpG, metilats o no, per tal d'afectar a l'expressió del gen.

6. CONCLUSIONS

Amb aquest treball s'han identificat dos gens, IGFBP1 i RN5S454, que podrien estar diferencialment expressats entre diabètics i no diabètics. El nivell d'evidència d'aquest resultat és molt limitat, tant per la magnitud de la diferència d'expressió (logFC) com pel nivell de significació estadística (p-valor). Per tant, cal ser prudents amb l'afirmació d'aquest resultat que requeriria una validació externa amb un estudi independent per tal de ser confirmat. A més a més s'han identificat dues seqüències d'origen desconegut que sembla ser que també es trobarien diferencialment expressades entre els dos grups. Tot i que hem intentat descobrir el seu origen, no sembla que els resultats que hem acabat obtenint ens l'expliquin.

Pel què fa el metiloma ens han sorgit varis dinucleòtids CpG diferencialment metilats, però entre aquests no hi hem trobat els dinucleòtids que pertanyen al gen IGFBP1, que es trobava diferencialment expressat entre diabètics i no diabètics.

No hem pogut realitzar una integració exhaustiva entre les dues òmiques, ja que els resultats eren molt limitats. Això podria ser degut a que el nostre model d'anàlisi és ajustat per obesitat, però els diabètics de tipus 2 tots ells eren obesos, a diferència dels no diabètics en els quals teníem tan obesos com no obesos. Un model d'anàlisi alternatiu hagués pogut ser una comparació entre els tres grups: diabètics de tipus 2 obesos, no diabètics obesos i no diabètics ni obesos.

7. PERSPECTIVES

A partir dels resultats d'aquest treball i des de la meua opinió personal, crec que s'hauria de considerar el fet de que el pseudogen ribosomal 454 i les seqüències desconegudes podrien estar implicades en la diabetis de tipus 2. De manera que podria ser interessant saber quin és el paper d'aquest pseudogen en les persones amb diabetis de tipus 2 i com es regula la seva expressió. Pel què respecta a les dues seqüències desconegudes crec que estaria bé poder consultar amb l'empresa que ha dissenyat el microarray i demanar més informació sobre les dues sondes.

8. BIBLIOGRAFIA

1. Nucleus Medical Media. (2016, 6 de juliol). *Diabetes tipo 1* [Vídeo]. Recuperat de <https://www.youtube.com/watch?v=fazeCkUgkwQ>
2. Nucleus Health Videos - Español. (2012, 17 d'octubre). *Diabetes de tipo 2* [Vídeo]. Recuperat de <https://www.youtube.com/watch?v=iIL-uC4f87U&t=157s>
3. Singh A., Gautier B., Shannon C.P., Vacher M., Rohart F., Tebbutt S.J., Cao K.L. (2016). DIABLO – an integrative, multi-omics, multivariate method for multi-group classification. *Bioinformatics*, 1-8. Doi: <https://doi.org/10.1101/067611>
4. Centro de Investigación Biomédica en Red de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), el Instituto de Salud Carlos III (Ministerio de Ciencia e Innovación), en colaboración con la Sociedad Española de Diabetes (SED) y la Federación Española de Diabetes (FED). (s.d.). La diabetes en España. Recuperat de <https://www.fundaciondiabetes.org/prensa/297/la-diabetes-en-espana>
5. Portal CLÍNIC. (2019). *Què és la diabetis de tipus 2?* Recuperat de <https://portal.hospitalclinic.org/ca/malalties/diabetis/diabetis-tipus-2#simptomes-1>
6. Ali, O. (2013). Genetics of type 2 diabetes. *World Journal of Diabetes*, 4(4), 114-123.
7. Ómica. (2017). Recuperat de <https://es.wikipedia.org/wiki/%C3%93mica>
8. Epigenómica. (2015). Recuperat de <https://www.genome.gov/es/about-genomics/fact-sheets/Epigenomica>
9. Cava C., Bertoli G., Castiglioni I. (2015). Integrating genetics and epigenetics in breast cancer: biological insights, experimental, computational methods and therapeutic potential. *BMC Systems Biology*, 9(62). Doi: <https://doi.org/10.1186/s12918-015-0211-x>
10. Methylome. (s.d.). Recuperat de <https://www.yourdictionary.com/methylome>
11. Long M.D., Smiraglia D.J., Campbell M.J. (2017). The Genomic Impact of DNA CpG Methylation on Gene Expression; Relationships in Prostate Cancer. *Biomolecules*, 7(1), 15.
12. Watson J.D., Baker T.A., Bell S.P., Gann A., Levine M., Losick R., Harrison S.C. (2014). *Biología Molecular del Gen* (7a ed.). New York: Pearson.

13. Hypermethylation. (2017). Recuperat de https://link.springer.com/referenceworkentry/10.1007/978-3-642-16483-5_2910
14. Hypomethylation. (2017). Recuperat de https://link.springer.com/referenceworkentry/10.1007/978-3-642-16483-5_2922
15. Nilsson E., Ling C. (2017). DNA methylation links genètics, fetal environment, and an unhealthy lifestyle to the development of type 2 diabetes. *Clinics Epigenetics*, 9, 105
16. Transcriptomics. (s.d.). Recuperat de <https://www.nature.com/subjects/transcriptomics>
17. Srivastava A., George J., Karuturi R.K.M. (2019). Transcriptome Analysis. *Encyclopedia of Bioinformatics and Computational Biology*, 3, 792-805.
18. Tracy Kim Kovach. (s.d). *Non-coding RNA (ncRNA)* [Videó]. Recuperat de <https://www.khanacademy.org/science/ap-biology/gene-expression-and-regulation/regulation-of-gene-expression-and-cell-specialization/v/non-coding-rna-ncrna>
19. Fang Y., Fullwood M.J. (2016). Roles, Functions and Mechanisms of Long Non-coding RNAs in Cancer. *Genomics, Proteomics & Bioinformatics*, 14(1), 42-54.
20. Xue M., Zhuo Y., Shan B. (2017). MicroRNAs, Long Noncoding RNAs, and Their Function in Human Disease. *Methods Mol Biol*, 1617, 1-25.
21. Dahlman-Wright K. (2016). *Series GSE64998*. Recuperat de <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE64998>
22. Dahlman-Wright K. (2019). *Series GSE65057*. Recuperat de <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi>
23. (2009). *Human Gene 1.1 ST Array Plate*. Recuperat de http://tools.thermofisher.com/content/sfs/manuals/human_gene_1_st_insert.pdf
24. GeneChip® Human Gene 1.1 ST Array Plate. (s.d.). Recuperat de <https://www.thermofisher.com/order/catalog/product/901415>
25. Fang Y., Fullwood M.J. (2016). Roles, Functions, and Mechanisms of Long Non-coding RNAs in Cancer. *Genomics, Proteomics & Bioinformatics*, 14(1). Doi: <https://doi.org/10.1016/j.gpb.2015.09.006>
26. Bolstad BM (2004). *Low Level Analysis of High-density Oligonucleotide Array Data: Background, Normalization and Summarization*. PhD thesis, University of California, Berkeley.

27. Bolstad BM, Collin F, Brettschneider J, Simpson K, Cope L, Irizarry RA, Speed TP (2005). “Quality Assessment of Affymetrix GeneChip Data.” In Gentleman R, Carey V, Huber W, Irizarry R, Dudoit S (eds.), *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, chapter 3, 33–47. Springer, New York.
28. Brettschneider J, Collin F, Bolstad BM, Speed TP (2007). “Quality assessment for short oligonucleotide arrays.” *Technometrics*, **In press**.
29. Ma plot. (2018). Recuperat de https://en.wikipedia.org/wiki/MA_plot
30. Bolstad, B.M., Irizarry R. A., Astrand M., and Speed, T.P. (2003), A Comparison of Normalization Methods for High Density Oligonucleotide Array Data Based on Bias and Variance. *Bioinformatics* 19(2):185-193;
31. Alboukadel Kassambara and Fabian Mundt (2017). factoextra: Extract and Visualize the Results of Multivariate Data Analyses.
R package version 1.0.5.
<https://CRAN.R-project.org/package=factoextra>
32. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015). “limma powers differential expression analyses for RNA-sequencing and microarray studies.” *Nucleic Acids Research*, 43(7), e47.
33. Gentleman R (2019). *annotate: Annotation for microarrays*. R package version 1.60.1.
34. MacDonald JW (2017). *hugene10sttranscriptcluster.db: Affymetrix hugene10 annotation data (chip hugene10sttranscriptcluster)*. R package version 8.7.0.
35. Bibikova M., Barnes B., Tsan C., Ho V., Klotzel B., Le J.M., Delano D., Zhang L., Schroth G.P., Gunderson K.L., Fan J.B., Shen R.(2011). High density DNA methylation array with single CpG site resolution. *Genomics*, 98(4). Doi: <https://doi.org/10.1016/j.ygeno.2011.07.007>
36. Weisenberger D.J., Van Den Berg D., Pan F., Berman B.P., Laird P.W. (2010). *Comprehensive DNA Methylation Analysis on the Illumina® Infinium® Assay Platform*. Recuperat de http://emea.support.illumina.com/content/dam/illumina-marketing/documents/products/appnotes/appnote_dna_methylation_analysis_infinium.pdf
37. Du P., Zhang X., Huang C.C., Jafari N., Kibbe W.A., Hou L., Lin S.M. (2010). Comparison of Beta-value and M-value methods for quantifying methylation

- levels by microarray analysis. *BMC Bioinformatics*, 11(587). Doi: <https://doi.org/10.1186/1471-2105-11-587>
38. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA (2014). “Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA Methylation microarrays.” *Bioinformatics*, **30**(10), 1363–1369. doi: [10.1093/bioinformatics/btu049](https://doi.org/10.1093/bioinformatics/btu049).
 39. Maksimovic J, Gordon L, Oshlack A (2012). “SWAN: Subset quantile Within-Array Normalization for Illumina Infinium HumanMethylation450 BeadChips.” *Genome Biology*, **13**(6), R44. doi: [10.1186/gb-2012-13-6-r44](https://doi.org/10.1186/gb-2012-13-6-r44).
 40. Fortin J, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, Greenwood CM, Hansen KD (2014). “Functional normalization of 450k methylation array data improves replication in large cancer studies.” *Genome Biology*, **15**(12), 503. doi: [10.1186/s13059-014-0503-2](https://doi.org/10.1186/s13059-014-0503-2).
 41. Triche TJ, Weisenberger DJ, Van Den Berg D, Laird PW, Siegmund KD (2013). “Low-level processing of Illumina Infinium DNA Methylation BeadArrays.” *Nucleic Acids Research*, **41**(7), e90. doi: [10.1093/nar/gkt090](https://doi.org/10.1093/nar/gkt090).
 42. Fortin J, Hansen KD (2015). “Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data.” *Genome Biology*, **16**, 180. doi: [10.1186/s13059-015-0741-y](https://doi.org/10.1186/s13059-015-0741-y).
 43. Andrews SV, Ladd-Acosta C, Feinberg AP, Hansen KD, Fallin MD (2016). “‘Gap hunting’ to characterize clustered probe signals in Illumina methylation array data.” *Epigenetics & Chromatin*, **9**, 56. doi: [10.1186/s13072-016-0107-z](https://doi.org/10.1186/s13072-016-0107-z).
 44. Fortin J, Triche TJ, Hansen KD (2017). “Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi.” *Bioinformatics*, **33**(4). doi: [10.1093/bioinformatics/btw691](https://doi.org/10.1093/bioinformatics/btw691)
 45. Fortin J, Fertig EJ, Hansen KD (2014). “shinyMethyl: interactive quality control of Illumina 450k DNA methylation arrays in R.” *F1000Research*, **3**, 175. doi: [10.12688/f1000research.4680.2](https://doi.org/10.12688/f1000research.4680.2).
 46. Hansen KD (2016). IlluminaHumanMethylation450kanno.ilmn12.hg19: Annotation for Illumina's 450k methylation arrays. R package version 0.6.0.
 47. Fold change. (2018). Recuperat de https://en.wikipedia.org/wiki/Fold_change

48. Factor de crecimiento insulínico tipo 1. (2015). Recuperat de https://es.wikipedia.org/wiki/Factor_de_crecimiento_insul%C3%ADnico_tipo_1#Mecanismo_de_acci%C3%B3n
49. UniProtKB – P05019 (IGF1_HUMAN). (2019). Recuperat de <https://www.uniprot.org/uniprot/P05019>
50. IGFBP1 Gene (Protein Coding). (s.d.). Recuperat de <https://www.genecards.org/cgi-bin/carddisp.pl?gene=IGFBP1&keywords=IGFBP1>
51. Petersson U., Ostgren C.J., Brudin L., Brsimar K., Nilsson P.M. (2009). Low levels of insulin-like growth-factor-binding protein-1 (IGFBP-1) are prospectively associated with the incidence of type 2 diabetes and impaired glucose tolerance (IGT): the Söderåkra Cardiovascular Risk Factor Study. *Diabetes Metab*, 35(3), 198-205.
52. Bae J.H., Song D.K., Im, S.S. (2013), Regulation of IGFBP-1 in Metabolic Diseases. *J Lifestyle Med*, 3(2), 73-79.
53. Random42 Scientific Communication. (2018, juny 21). IGF Oncology [Vídeo]. Recuperat de <https://www.youtube.com/watch?v=b5hBWssh4aE>
54. RNA5SP454 RNA, 5S ribosomal pseudogene 454 [Homo sapiens (human)]. (2019). Recuperat de <https://www.ncbi.nlm.nih.gov/gene/100873699>
55. Kumar M, Nath S, Prasad HK, Sharma GD, Li Y. (2012). *Protein Cell*, 3(10), 726-738.
56. Insulin receptor. (2019) Recuperat de https://en.wikipedia.org/wiki/Insulin_receptor
57. Insulin signal transduction pathway. (2019). Recuperat de https://en.wikipedia.org/wiki/Insulin_signal_transduction_pathway
58. Details for Transcript Cluster 7894124. (s.d.). Recuperat de https://www.affymetrix.com/analysis/netaffx/exon/wtgene_transcript.affx?pk=4:7894124
59. Details for Trasncript Cluster 7895099. (s.d.). Recuperat de https://www.affymetrix.com/analysis/netaffx/exon/wtgene_transcript.affx?pk=4:7895099
60. Question: Control probe sets in Affymetric ST microarray. (2016). Recuperat de <https://www.biostars.org/p/167209/>

61. Exon and Gene Array Glossary. (s.d.). Recuperat de http://www.affymetrix.com/support/help/exon_glossary/index.affx
62. DNA microarray. (2019). Recuperat de https://en.wikipedia.org/wiki/DNA_microarray
63. Kozomara A, Birgaoanu M, Griffiths-Jones S. (2019). miRBase: from microRNA sequences to function. *Nucleic Acids Res*, 47, 155-162.
64. Kozomara A, Griffiths-Jones S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res*, 42, 68-73.
65. Kozomara A, Griffiths-Jones S. (2011). miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acid Res*, 39, 152-157.
66. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. (2008). miRBase: tools for microRNA genomics. *Nucleic Acid Res*, 36, 154-158.
67. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. (2006). miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acid Res*, 34, 140-144.
68. Griffiths-Jones S. (2004). The microRNA Registry. *Nucleic Acid Res*, 32, 109-111.
69. Frequently Asked questions. (s.d.). Recuperat de https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=FAQ#expect
70. Symbol report for MIR3916. (s.d.). Recuperat de https://www.genenames.org/data/gene-symbol-report#!/hgnc_id/HGNC:38912
71. Quek XC, Thomson DW, Maag JL, Bartonicek N, Signal B, Clark MB, Gloss BS and Dinger ME (2014) lncRNADB v2.0: expanding the reference database for functional long noncoding RNAs. *Nucleic Acid Res* 43: D168-D173
72. Amaral PP, Clark MB, Gascoigne DK, Dinger ME, Mattick JS (2011). lncRNADB: a reference database for long noncoding RNAs. *Nucleic Acids Res* 39: D146-151.
73. Batch effects. (s.d.). Recuperat de http://www.molmine.com/magma/global_analysis/batch_effect.html
74. FPR2 Gene (Protein Coding). (s.d.). Recuperat de <https://www.genecards.org/cgi-bin/carddisp.pl?gene=FPR2&keywords=FPR2>
75. Serrano CJ, Cuevas-Córdoba B, Macías-Segura N, González-Curiel RA, Martínez-Balderas VY, Enciso-Moreno L, Small P, Hernández-Pando R, Enciso-

Moreno JA. (2016). Transcriptional profiles discriminate patients with pulmonary tuberculosis from non-tuberculous individuals depending on the presence of non-insulin diabetes mellitus. *Clin Immunol*, 162, 107-117.

76. Akt/PKB signaling pathway. (2018). Recuperat de https://en.wikipedia.org/wiki/Akt/PKB_signaling_pathway
77. EHD4 Gene (Protein Coding). (s.d.). Recuperat de <https://www.genecards.org/cgi-bin/carddisp.pl?gene=EHD4&keywords=EHD4>
78. SAMD11 Gene (Protein Coding). (s.d.). Recuperat de <https://www.genecards.org/cgi-bin/carddisp.pl?gene=SAMD11&keywords=SAMD11>
79. SPOCK1 Gene (Protein Coding). (s.d.). Recuperat de <https://www.genecards.org/cgi-bin/carddisp.pl?gene=SPOCK1&keywords=SPOCK1>
80. KRTAP24-1 Gene (Protein Coding). (s.d.). Recuperat de https://www.genecards.org/cgi-bin/carddisp.pl?gene=KRTAP24-1&keywords=KRTAP24-1#pathways_interactions
81. ZBTB22 Gene (Protein Coding). (s.d.). Recuperat de <https://www.genecards.org/cgi-bin/carddisp.pl?gene=ZBTB22&keywords=ZBTB22>
82. ARFGEF3 Gene (Protein Coding). (s.d.). Recuperat de <https://www.genecards.org/cgi-bin/carddisp.pl?gene=ARFGEF3&keywords=ARFGEF3>
83. Metablism of proteins. Super Path. (2019). Recuperat de https://pathcards.genecards.org/card/metabolism_of_proteins
84. TIMP4 Gene (Protein Coding). (s.d.). Recuperat de <https://www.genecards.org/cgi-bin/carddisp.pl?gene=TIMP4&keywords=TIMP4>
85. SYN2 Gene (Protein Coding). (s.d.). Recuperat de <https://www.genecards.org/cgi-bin/carddisp.pl?gene=SYN2&keywords=SYN2>
86. NSD1 Gene (Protein Coding). (s.d.). Recuperat de <https://www.genecards.org/cgi-bin/carddisp.pl?gene=NSD1&keywords=NSD1>
87. HID1 Gene (Protein Coding). (s.d.). Recuperat de <https://www.genecards.org/cgi-bin/carddisp.pl?gene=HID1&keywords=C17orf28>
88. CpG sites. (2019). Recuperat de https://en.wikipedia.org/wiki/CpG_site

ANNEX

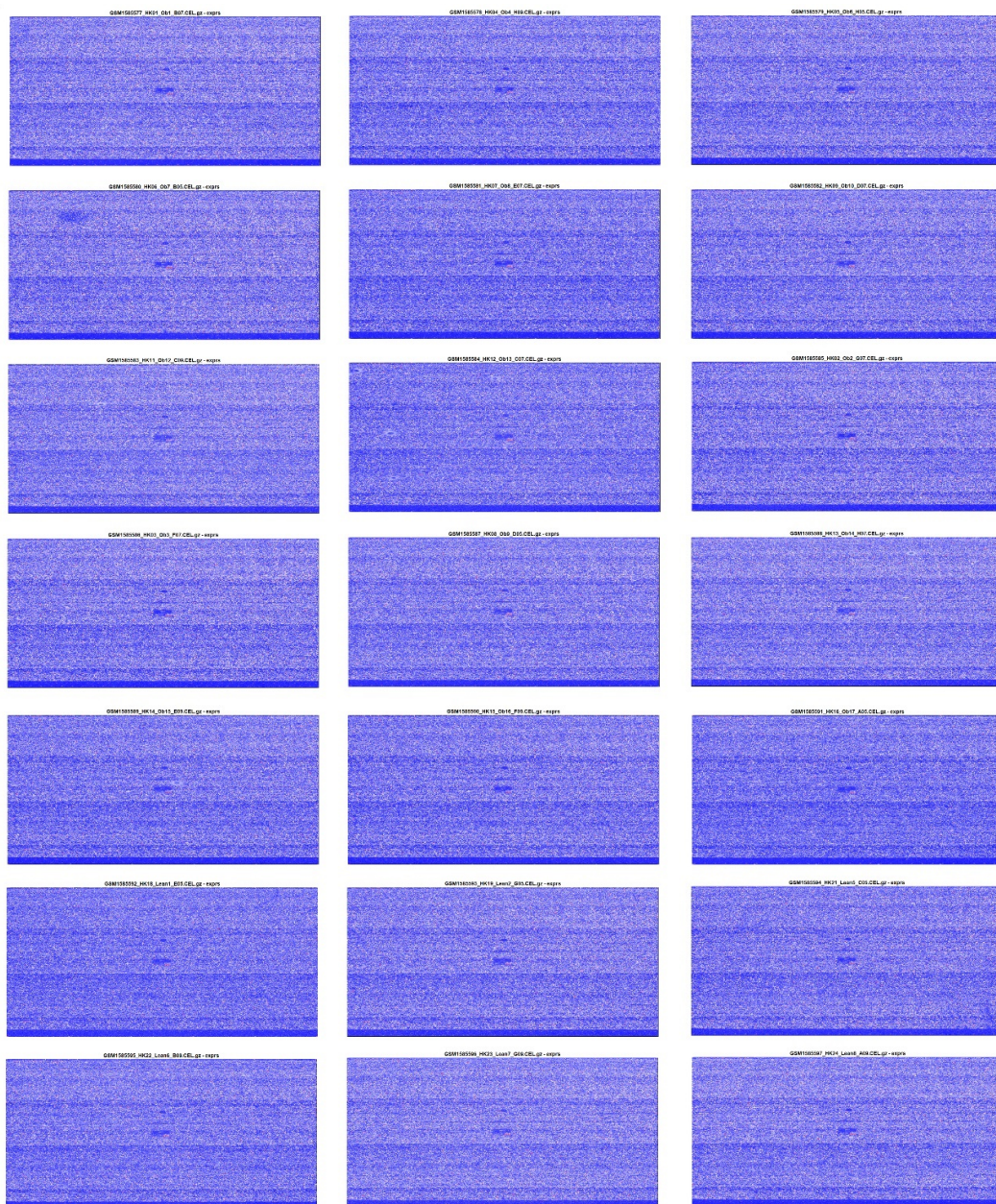


Figura A1: Imatge dels 21 arrays corresponents a les mostres del transcriptoma. No s'observa cap desplaçament de les sondes, ni contaminants; els arrays estan intactes.

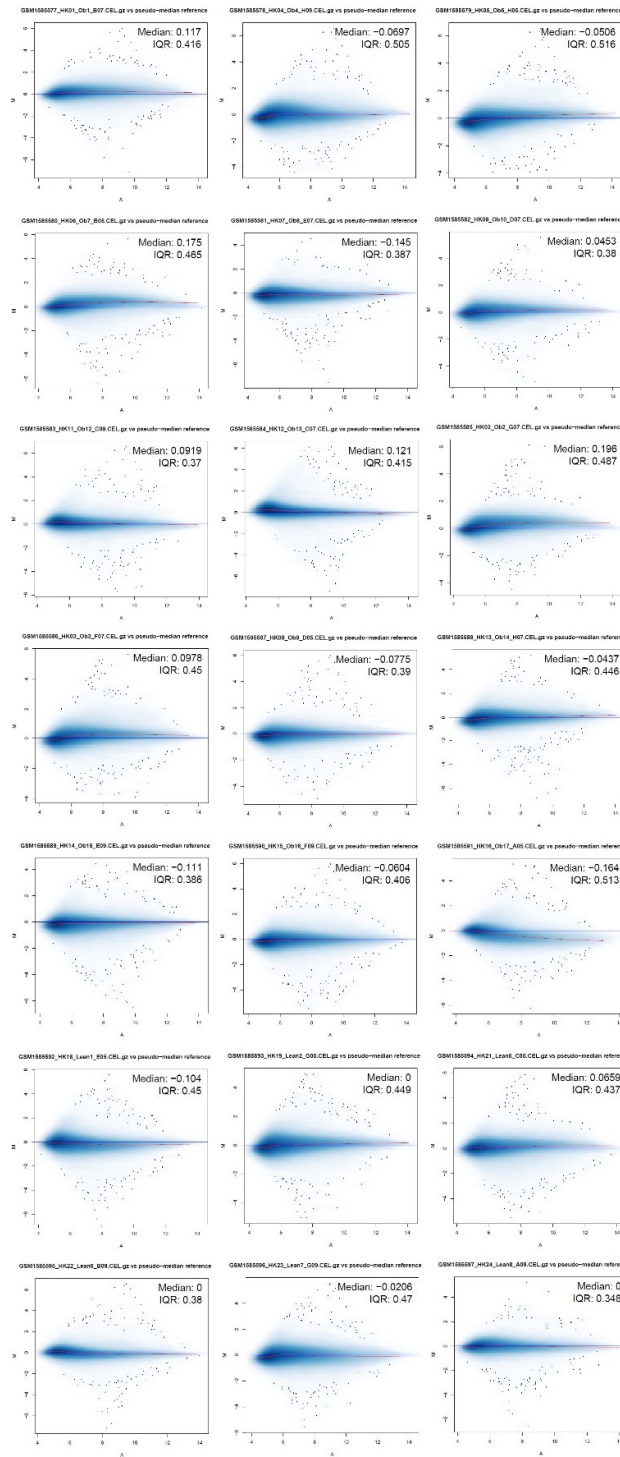


Figura A2: MAplot de les mostres del transcriptoma no normalitzades. En l'eix X hi trobem el logartime de la mediana d'intensitat de les mostres i en l'eix Y el logartime de la distribució de les intensitats