Metilació d'ADN en càncer de pàncrees: el fenotip metilador d'illes CpG i el seu efecte en la supervivència

Ferran Juanals Figueras
Màster en Bioinformàtica i Bioestadística
Anàlisis de dades òmiques

Jeroni Luna Cornadó
David Merino Arranz

01/2019

Aquesta obra està subjecta a una llicència de Reconeixement-NoComercial-SenseObraDerivada 3.0 Espanya de Creative Commons

FITXA DEL TREBALL FINAL

Títol del treball:	*Metilació d'ADN en càncer de pàncrees: el fenotip metilador d'illes CpG i el seu efecte en la supervivència*
Nom de l'autor:	*Ferran Juanals Figueras*
Nom del consultor/a:	*Jeroni Luna Cornadó*
Nom del PRA:	*David Merino Arranz*
Data de lliurament (mm/aaaa):	*01/2019*
Titulació o programa:	*Màster en Bioinformàtica i Bioestadística*
Àrea del Treball Final:	*Anàlisis de dades òmiques*
Idioma del treball:	*Català*
Paraules clau	*metilació d'ADN, CIMP, supervivència*

Resum del Treball (màxim 250 paraules):

S'han analitzat les dades clíniques i de metilació de 184 càncers de pàncrees obtingudes de *The Cancer Genome Atlas* (TCGA) per observar les possibles relacions entre el fenotip metilador d'illes CpG (CIMP) i la supervivència. Les mostres han estat classificades mitjançant un l'algoritme K-means i s'ha trobat que 50 tenien alts valors mitjans de metilació (CIMPH) i 56 els tenien baixos (CIMPL). S'ha analitzat les corbes de supervivència Kaplan-Meier d'aquests dos grups i s'ha vist que existeixen diferències significatives entre els dos grups, indicant que l'hipermetilació aberrant representativa de CIMPH està relacionada amb una disminució de la supervivència dels pacients de càncer de pàncrees. Un anàlisis de metilació diferencial ha permès identificar 4916 sondes amb alts nivells de metilació que corresponen a 1423 gens candidats a ser marcadors per al diagnòstic de càncer de pàncrees. Aquest treball pot ajudar a definir més clarament els subgrups CIMP dintre del càncer de pàncrees i la seva relació amb la supervivència.

Abstract (in English, 250 words or less):

We analysed clinical and methylation data from 184 pancreatic tumor samples from The Cancer Genome Atlas (TCGA) to assess the relationship between CpG island methylator phenotype (CIMP) and the survival. A classification with the k-means algorithm found 50 tumor samples with higher average levels of methylation (CIMPH) and 56 tumor samples with lower average levels of methylation (CIMPL). We analysed the survival curve for the both groups with the Kaplan-Meier estimate and found that they have different survival functions, having CIMPH worse survival. A differential methylation analysis identified 4916 CpG sites with aberrant hypermetilation, 1423 candidate genes to be markers for diagnostic of pancreatic cancer. These results can help to refine existing CIMP subtypes of pancreatic cancer and its relation with survival.

Índex

1. Introducció 1
1.1 Context i justificació del Treball 1
1.2 Objectius del Treball 1
1.3 Enfocament i mètode seguit 2
1.4 Planificació del Treball 2
1.5 Breu sumari de productes obtinguts 3
1.6 Breu descripció dels altres capítols de la memòria 3
2. Materials i mètodes 4
2.1 Dades biològiques 4
2.3 Mètodes estadístics 5
3. Resultats 6
3.1 Classificació de les mostres tumorals en els subtipus CIMP 6
3.2 Anàlisi de supervivència 8
3.3 Anàlisi de regions diferencialment metilades 9
3.4 Anàlisi de significació biològica 9
4. Conclusions 10
5. Glossari 12
6. Bibliografia 13
7. Annexos 15

Llista de figures

Figura 1. Diagrama de Gantt que il·lustra la planificació del projecte..... 2
Figura 2. Gràfic box-plot de les mitjanes de metilació d'ADN de les mostres tumorals agrupades en els tres grups CIMP 7
Figura 3. Gràfic de les corbes de supervivència Kaplan-Meier per als grups CIMPH i CIMPL 8
Figura 4. Volcano plot de les diferències entre les mitjanes de metilació de CIMPH i CIMPL 9
Figura 5. Gràfic de barres per a la visualització dels principals termes de l'anàlisi de significació biològica 10

1. Introducció

1.1 Context i justificació del treball

El càncer de pàncrees és un dels càncers amb un índex de mortalitat més elevat degut a la manca de marcadors específics per al diagnòstic durant les etapes inicials, la gran agressivitat biològica de la malaltia i la resistència als medicaments citotòxics. Aquests factors provoquen que els pacients de càncer de pàncrees tinguin una supervivència mitjana de 6 mesos i una taxa de supervivència del 6% als 5 anys[1].

Els canvis epigenètics juguen un paper molt important en la iniciació i progressió dels tumors. Entre aquest canvis s'hi troba la metilació de l'ADN, que té un paper fonamental en la regulació de l'expressió gènica[2].

S'ha descrit l'existència de nivells elevats de metilació de l'ADN de manera amplia en diferents tipus de càncer, el que s'ha anomenat com fenotip metilador d'illes CpG (CIMP)[3]. CIMP ha estat descrit com un fenomen que pot ajudar a la predicció de la prognosis i la resposta a tractaments en diferents càncers, però encara no està clar si és un fenomen global dels processos tumorals o si està lligat específicament a cada tipologia de càncer[3].

CIMP va ser identificat primer en càncer colorectal per *Toyota et al.*[4], però des d'aleshores s'ha descrit en gran varietat de tipologies de tumor, incloent-hi el de pàncrees[5].

La informació continguda en The Cancer Genome Atlas (TCGA) ens dona la possibilitat d'estudiar les dades de metilació de càncer de pàncrees a gran escala per a intentar descriure els grups CIMP i analitzar les seves diferencies en relació a la supervivència.

1.2 Objectius del treball

- Obtenció de les dades clíniques i de metilació de càncer de pàncrees del projecte TCGA.
- Classificació de les mostres tumorals en grups representatius del nivell CIMP.
- Anàlisi de supervivència dels diferents grups CIMP.
- Identificar els gens amb diferencies de metilació entre els grups CIMP.
- Anàlisi de significació biològica dels gens identificats.

1.3 Enfocament i mètode seguit

L'estudi està enfocat a l'anàlisi *in silico* de les dades clíniques i de metilació en relació amb la supervivència de les mostres tumorals de càncer de pàncrees obtingudes del consorci internacional TCGA.

Les dades recollides pel projecte TCGA que s'utilitzaran per a l'anàlisi s'obtindran del portal Genomic Data Commons Data Portal (GDC Data Portal), mantingut per el National Cancer Institute (NCI) d'Estats Units. D'entre totes les dades disponibles, es descarregaran les corresponents als anàlisis de metilació i a les dades clíniques.

A partir d'aquestes dades es pretén identificar les diferències en els patrons de metilació dels pacients amb diferents temps de supervivència a la malaltia. Un cop s'identifiquin els gens afectats per els diferents nivells de metilació, es realitzarà un anàlisis de significació biològica.

1.4 Planificació del treball

- Obtenció de les dades de les anàlisis clíniques i de metilació del càncer de pàncrees del dipòsit online i lectura de bibliografia.
- Creació i processament d'una nova base de dades a partir de les dues obtingudes del dipòsit online.
- Identificar les sondes de metilació que presentin diferencies significatives entre els nivells de supervivència.
- Identificar els gens corresponents a les sondes de metilació amb diferencies significatives entre els nivells de supervivència.
- Determinar un patró que caracteritzi els diferents nivells de supervivència a partir de les diferencies en el nivell de metilació.
- Anàlisis de significació biològica dels patrons dels diferents nivells de supervivència.
- Avaluació i cerca bibliogràfica de les implicacions biològiques dels resultats obtinguts.

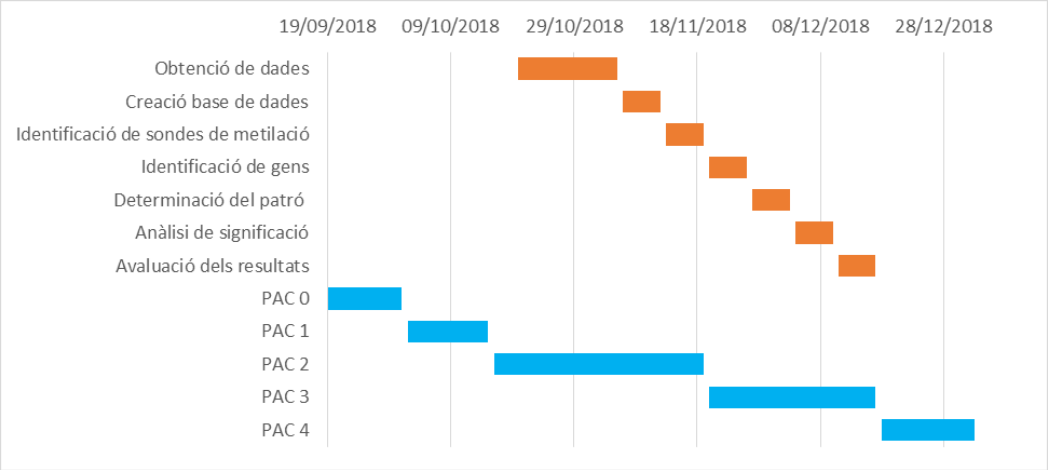

Figura 1. Diagrama de Gantt que il·lustra la planificació del projecte

1.5 Breu sumari de productes obtinguts

- Classificació de les mostres tumorals de càncer de pàncrees incloses en el projecte TCGA-PAAD en tres grup representatius dels fenotips metiladors de illes CpG.
- Resultats de l'anàlisi de supervivència de Kaplan-Meier entre CIMPH i CIMPL.
- Llistat de gens corresponents a les sondes hipermetilades en CIMPH respecte a CIMPL.
- Resultats de l'anàlisi de significació biològica per els termes de Gene Ontology.

1.6 Breu descripció dels altres capítols de la memòria

En el segon capítol, Materials i mètodes, es detallen els materials utilitzats en l'anàlisi des de l'obtenció de les dades, els softwares utilitzats fins als mètodes estadístics aplicats.

A continuació en el tercer capítol, Resultats, es documenta en diferents seccions els resultats obtinguts en els diferents anàlisis del projecte, les raons per les quals s'ha realitzat, les possibles alternatives o variacions i la discussió dels resultats en base a les evidencies científiques actuals.

Finalment, en el quart capítol, es detallen les conclusions obtingudes de l'estudi, una crítica a la planificació del projecte i una secció d'aspectes futurs en relació a l'estudi realitzat.

2. Materials i mètodes

2.1 Dades biològiques

El projecte es desenvolupa a partir de les dades biològiques de 184 mostres tumorals de càncer de pàncrees obtingudes del projecte TCGA. D'aquestes mostres se'n obtindran les dades clíniques dels pacients i els perfils de metilació de l'ADN.

El TCGA és un projecte finançat públicament i supervisat per el National Cancer Institute (NCI) i el National Human Genome Research Institute (NHGRI) d'Estats Units que pretén catalogar i descobrir les alteracions genòmiques causants dels càncers per crear un atles que descrigui els perfils genòmics de cada un d'ells. A dia d'avui, els investigadors de TCGA han analitzat més de 30 tipus de tumors humans gracies a l'aplicació de tecnologies de seqüenciació de nova generació[6,7].

Un dels objectius principals del projecte TCGA és que els conjunts de dades estudiats estiguin a disposició de la comunitat investigadora, per això totes les dades son accessibles des del portal Genomic Data Commons Data Portal (GDC Data Portal). És des d'aquest portal d'on obtindrem les dades clíniques i de metilació d'ADN corresponents a càncer de pàncrees, englobades dins el projecte TCGA-PAAD. El conjunt de dades del projecte TCGA-PAAD consta de 185 casos, dels quals 184 corresponen a mostres provinents de tumor primari que seran les utilitzades per als anàlisis.

Les dades de metilació obtingudes del GDC Data Portal provenen de l'anàlisi de la plataforma HumanMethylation450K de Illumina, consten de 184 mostres de tumor primari i 10 casos control. Les dades han passat per el pre-processament de nivell 3 de TCGA. Les sondes localitzades en els cromosomes X i Y han estat descartades, així com les sondes que no presenten valors per algun dels casos.

Les dades clíniques també han estat obtingudes del GDC Data Portal i consten de 184 casos.

2.2 Software pel desenvolupament del treball

Per a l'obtenció i l'anàlisi de les dades del treball s'ha utilitzat majoritàriament el paquet TCGAbiolinks[8] dins del software lliure R/Bioconductor[9,10], mitjançant la interfície RStudio. El paquet TCGAbiolinks proporciona les eines necessàries per a l'obtenció, el processament i l'anàlisi de las dades del GDC Data Portal. Per a l'anàlisi de significació biològica s'utilitzarà la base de dades Gene Ontology[11], que recull el coneixement de les funcions dels gens i els seus productes.

2.3 Mètodes estadístics

2.3.1 Classificació de les mostres tumorals en els subtipus CIMP

El primer pas consisteix en seleccionar un conjunt de sondes CpG representatives per a l'estudi, tal i com es descriu en l'article de *Sánchez-Vega F et al.*[12]. Del conjunt de dades de metilació de nivell 3 obtingut, s'examinaran les sondes localitzades en illes CpG i s'exclouran les localitzades en els cromosomes X i Y, les sondes en què manca algun valor per algun dels casos i les sondes amb molt baixa variabilitat (SD<0'1). La selecció de sondes es realitzarà tenint en compte dos paràmetres representatius de la diferència de metilació entre les mostres tumorals i les mostres control. Una sonda serà seleccionada per a la inclusió en el grup representatiu de metilació diferenciada en cas que el seu nivell mitjà de metilació sobre les mostres control sigui inferior a 0'05 i el seu nivell mitjà de metilació sobre les mostres tumorals sigui superior a 0'25.

La classificació de les mostres en els grups representatius de l'estat CIMP es realitzarà amb l'algoritme K-means, agrupant els valors mitjans de metilació de les mostres calculats sobre el conjunt de sondes diferencialment metilades.

2.3.2 Anàlisi de supervivència

L'anàlisi de supervivència es realitzarà amb la funció TCGAanalyze_survival del paquet TCGAbiolinks[8] que utilitza l'estimador de Kaplan-Meier i en representa les corbes pels dos grups. L'estimador Kaplan-Meier permet la presencia de dades censurades, que es el cas per algunes de les mostres on s'ha perdut el seguiment. Per a comparar les dues corbes de Kaplan-Meier s'utilitza la proba de Mantel-Cox, on la hipòtesi nul·la assumeix que els dos grups tenen la mateixa funció de supervivència[13].

2.3.3 Anàlisi de regions diferencialment metilades

S'analitzarà les sondes CpG diferencialment metilades entre els dos grups CIMP amb la funció TCGAanalyze_DMR del paquet TCGAbiolinks[8]. Aquesta funció calcula la diferencia entre les mitjanes de metilació de cada grup per a cada sonda i les compara amb el test de Wilcoxon ajustant pel mètode de Benjamini-Hochberg. Perquè una sonda sigui considerada diferencialment metilada ha de presentar una diferència de metilació absoluta major a 0'3 i un p-valor ajustat menor a 10^{-5}.

La funció també crea un Volcano Plot on l'eix de les X representa la diferencia de metilació mitjana i l'eix de les Y representa la significació.

2.3.4 Anàlisi de significació biològica

S'utilitzarà les funcions TCGAanalyze_EAcomplete per a la realització de l'anàlisi de significació biològica de Pathway i Gene Ontology[11] (Biological Process, Cellular Component i Molecular Function). Amb la funció TCGAvisualize_EAbarplot del paquet TCGAbiolinks[8] es representarà un gràfic de barres on hi apareixen els termes GO més representatius dels gens analitzats.

3. Resultats

3.1 Classificació de les mostres tumorals en els subtipus CIMP

Per realitzar la classificació de les mostres tumorals s'analitzen les dades de metilació d'ADN de 184 casos de tumor primari de càncer de pàncrees i 10 casos control obtingudes del projecte TCGA a través del GDC Data Portal. Les dades provenen de la plataforma HumanMethylation450K d'Illumina i contenen els valors beta de més de 485.000 sondes per a cada una de les mostres i controls. Degut a que el fenotip metilador CIMP ha estat descrit en illes de metilació CpG, seran les sondes localitzades en aquestes illes que s'examinaran en les conseqüents anàlisis. De totes les sondes localitzades en illes de metilació CpG, s'exclouran les que estiguin localitzades en els cromosomes X o Y, les que no presentin valors per a totes les mostres i les que tinguin metilació amb baixa variabilitat (SD<0'1) entre les mostres tumorals[12].

Així dons, l'examinació de les 31.865 sondes que presenten variació ens permetrà seleccionar-ne un grup representatiu, que compleixi el patró de metilació diferenciada on les sondes presenten un grau de metilació mínim en les mostres control i un augment en el grau de metilació en les mostres tumorals. La selecció representarà les sondes amb nivells de metilació mitjans per sota del 5% en les mostres control i per sobre del 25% en les mostres tumorals[12]. El número de sondes diferencialment metilades seleccionades per a la classificació de les mostres tumorals en els subtipus CIMP és de 750.

L'obtenció dels diferents grups representatius de l'estat CIMP s'ha realitzat mitjançant la classificació de les mostres tumorals en tres categories amb l'ús d'un algoritme K-means, agrupant els valors mitjans de metilació de les mostres calculats sobre el conjunt de sondes diferencialment metilades. El grup de mostres amb els valors mitjans més baixos de metilació s'ha anomenat CIMPL (n=56) i el grup amb els valors mitjans més alts CIMPH (n=50), aquests son els dos grups que s'utilitzaran per a les posteriors anàlisis proposades en aquest treball. El tercer grup de mostres, amb uns valors mitjans de metilació intermedis,

s'ha anomenat CIMPi. Les mostres classificades com a CIMPi s'han exclòs de les posteriors anàlisis, d'aquesta manera les mostres tumorals comparades presenten unes diferències biològiques més substancials i son més representatives de l'estat CIMP.

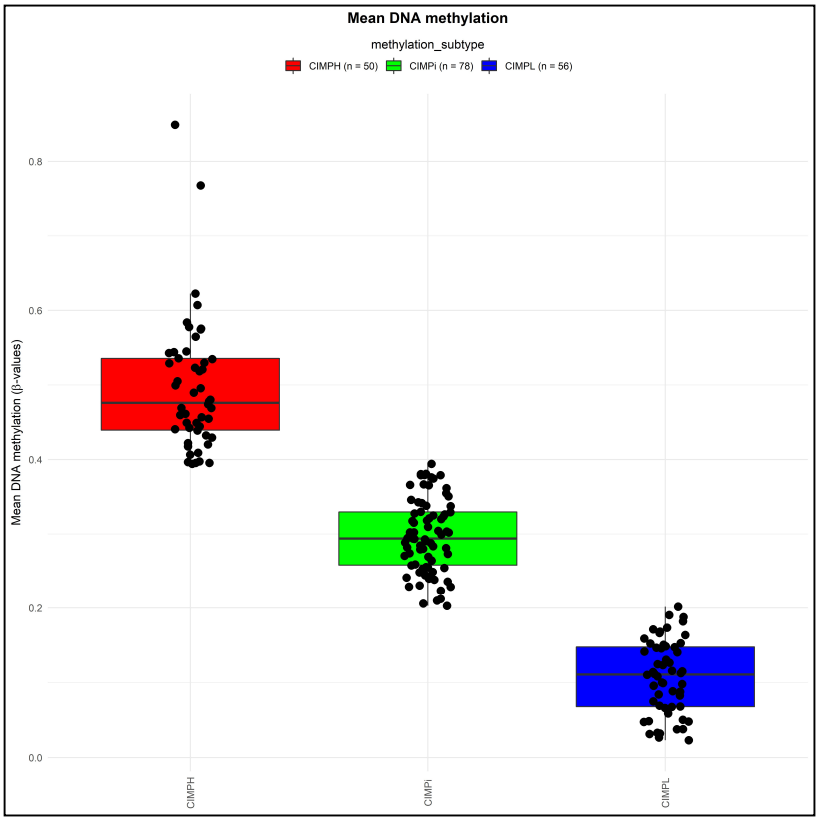

Figura 2. Gràfic box-plot de les mitjanes de metilació d'ADN de les mostres tumorals agrupades en els tres grups CIMP

En el gràfic box-plot de la Figura 2 es pot observar les diferències entre els valors mitjans de metilació de les diferents mostres agrupades per la categoria CIMP assignada. Es pot apreciar com els valors mitjans del grup intermedi CIMPi formen un continu entre les mostres dels grups CIMPH i CIMPL, per tant, les diferències entre aquests i el grup CIMPi son inexistents per algunes mostres. L'exclusió de CIMPi permetrà la comparació de dos grups inequívocament diferents, de manera que s'estudiaran les mostres representatives dels extrems CIMP. Tanmateix, essent CIMPi el grup més nombrós (n=78), l'exclusió reduirà la mida mostral de les següents anàlisis estadístiques substancialment.

Un altre dels aspectes a tenir en compte és l'elecció de l'algoritme K-means per a la classificació de les mostres. La senzillesa i eficiència de K-means a l'hora de classificar les mostres juga al seu favor, però l'algoritme també té els seus inconvenients. Una de les possibles millores del treball podria ser l'avaluació d'altres algoritmes per classificar les mostres.

3.2 Anàlisi de supervivència

Un cop definits els dos grups CIMP d'interès, es comprovarà si l'amplia hipermetilació de les mostres CIMPH i la hipometilació de les mostres CIMPL provoquen diferencies en les respectives funcions de supervivència. S'utilitza l'estimador de Kaplan-Meier per a representar les funcions de supervivència (Figura 3). L'estimador de Kaplan-Meier té en compte els casos censurats, presents en les dades clíniques obtingudes a través del GDC Data Portal. El conjunt de dades clíniques conté 43 variables per a les 184 mostres tumorals. Entre aquestes variables hi trobem l'estat vital del pacient (85 vius; 99 morts), els dies transcorreguts fins la seva mort i els dies transcorreguts des de l'última visita, entre d'altres.

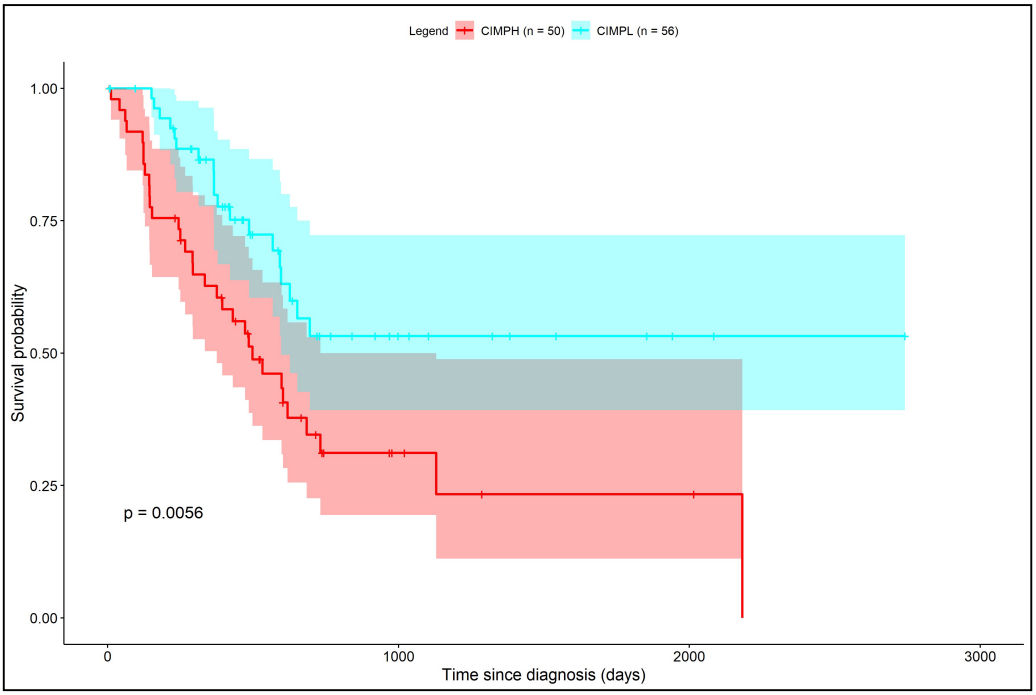

Figura 3. Gràfic de les corbes de supervivència Kaplan-Meier per als grups CIMPH i CIMPL

En el gràfic de la Figura 3 estan representades les dues corbes de Kaplan-Meier de les mostres representatives de CIMPH i CIMPL, i es pot observar que les mostres CIMPH tenen una probabilitat de supervivència més baixa que les CIMPL. La proba de Mantel-Cox, amb un p-valor de 0'006, ens indica que si que existeixen diferències significatives entre la supervivència dels dos grups.

Tal i com s'ha comentat a l'apartat anterior, el fet d'haver exclòs CIMPi de l'anàlisi de supervivència ha reduït quasi a la meitat el nombre de mostres, i per tant, la introducció de les mostres classificades en aquest grup pot fer variar el resultat obtingut. A part d'això, també cal tenir en compte que tant l'estimador Kaplan-Meier com la proba de Mantel-Cox assumeixen que la censura no té relació amb la prognosis i aquest pot no ser el cas[13].

3.3 Anàlisi de regions diferencialment metilades

En aquest apartat es vol conèixer quines sondes presenten un nivell de metilació significativament diferent entre el grup CIMPH i el grup CIMPL. Es busca determinar quins gens es troben hipermetilats en les mostres CIMPH respecte CIMPL per a proposar-los com a bio-marcadors en càncer de pàncrees.

Per a l'anàlisi s'utilitzarà el conjunt de sondes que presenten variació en la metilació, 31.865 sondes localitzades en illes CpG. Per a cada sonda es calcularà la diferència entre les mitjanes dels dos grups i es realitzarà la proba de Wilcoxon ajustant pel mètode de Benjamini-Hochberg. Les sondes seran considerades diferencialment metilades en el cas que la diferència de metilació entre les mitjanes dels dos grups tingui un valor absolut igual o superior a 0'3 i el p-valor ajustat de la proba de Wilcoxon sigui menor a 10^{-5}.

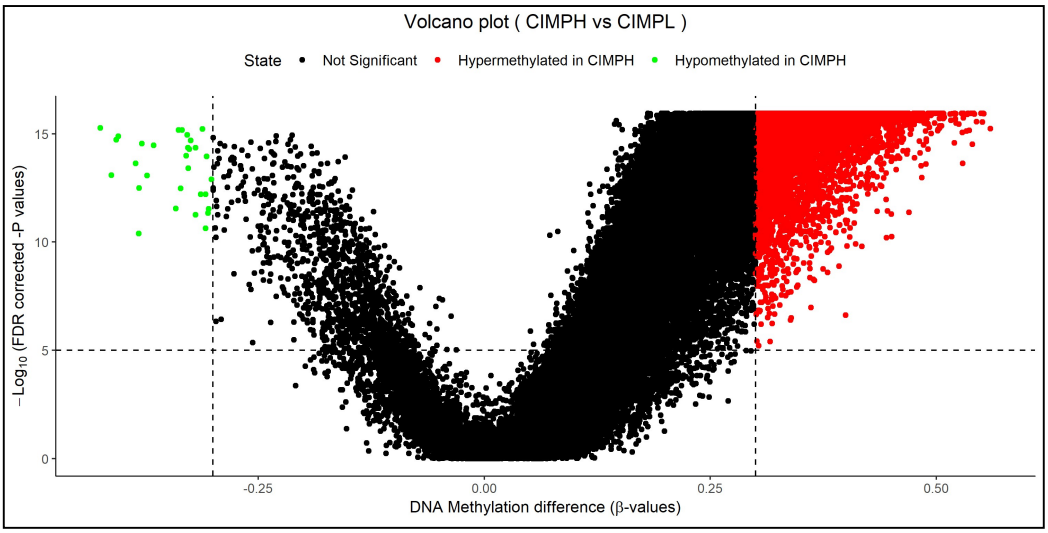

Figura 4. Volcano plot de les diferències entre les mitjanes de metilació de CIMPH i CIMPL

En el "Volcano plot" de la Figura 4 es poden observar les sondes hipermetilades en el grup CIMPH respecte a CIMPL representades en color vermell (4916 sondes) i les sondes hipometilades en color verd (30 sondes). Les 4916 sondes hipermetilades corresponen a 1423 gens diferents, amb els quals es realitzarà un anàlisi de significació biològica (el llistat dels gens hipermetilats en les mostres CIMPH es pot trobar als annexos).

3.4 Anàlisi de significació biològica

En aquest últim apartat es realitzarà un anàlisi de significació biològica del grup de gens hipermetilats en les mostres CIMPH. L'anàlisi ens permetrà descobrir els termes GO que estan sobre-representats en el

grup de gens seleccionats. En la Figura 5 es mostren els termes GO per funcions moleculars (MP), components cel·lulars (CC), processos biològics (BP) i rutes de senyalització (Pathway) més representatius per als gens analitzats.

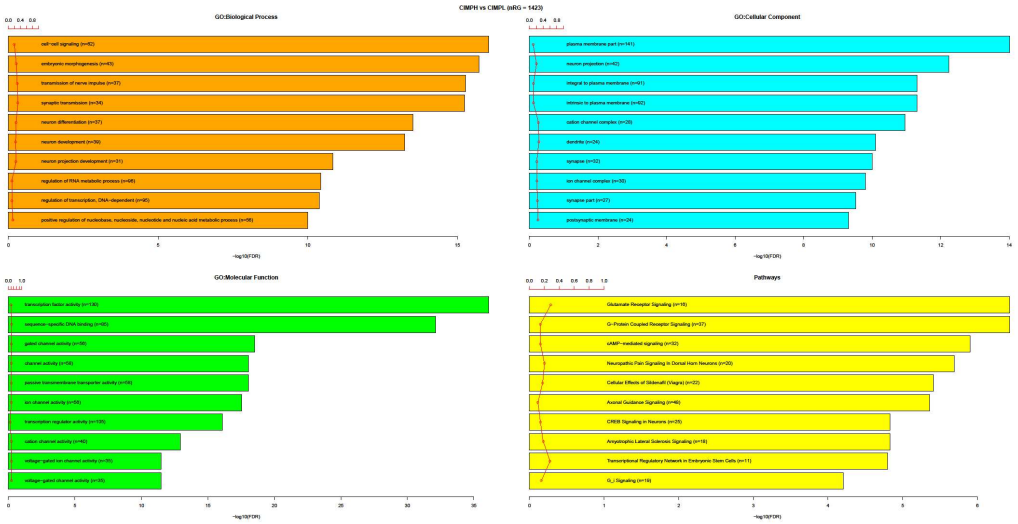

Figura 5. Gràfic de barres per a la visualització dels principals termes de l'anàlisi de significació biològica.

4. Conclusions

La identificació de mostres CIMP en tumors de pàncrees pot ser de gran importància per al diagnòstic i tractament d'aquest càncer. Tal i com s'ha vist, el fenotip de hipermetilació aberrant de CIMPH té afecte negatiu a la supervivència dels pacients de càncer de pàncrees. Per a identificar els pacients representatius de CIMPH és important establir un patró de gens que inequívocament determinin aquest fenotip hipermetilat. El llistat de gens obtingut és un primer pas per a l'assoliment d'aquest objectiu, però seria necessari un estudi més profund dels gens inclosos en el llistat, així com analitzar la relació entre aquests gens altament metilats i els seus nivells d'expressió. Així dons, encara que s'han assolit els objectius plantejats, aquest treball deixa molt de marge d'aprofundiment i millora.

Encara que en un primer moment no estava previst la classificació de les mostres en categories CIMP, es va canviar l'enfoc del treball per incloure-les, passant a ser una part important del treball. Aquesta decisió es va donar després de la lectura de bibliografia on s'exposava les possibilitats d'aquest tipus de classificació.

Pel que fa a la planificació, s'ha de reconèixer que no ha estat la correcta i que no s'ha dedicat al treball el temps necessari ni de bon tros. Des d'un primer moment, hi ha hagut un endarreriment en la realització i entrega de les diferents proves d'avaluació contínua que s'ha anat

incrementant en el temps i que ha acabat provocant que el treball entregat no estigui realitzat de la millor manera possible.

Tal i com s'ha comentat anteriorment, aquest treball té diversos punts de possible millora, com pot ser buscar i analitzar altres mètodes de selecció i classificació de les sondes en grups CIMP. Un altre aspecte important que no s'ha pogut explorar en aquest treball és la integració dels valors d'expressió gènica en les anàlisis, quelcom necessari a l'hora de seleccionar un patró de gens representatius de hipermetilació en CIMP.

5. Glossari

En ordre alfabètic:

ADN: Àcid desoxiribonucleic

CIMP: Fenotip metilador d'illes CpG

CpG: 5'—C—phosphate—G—3'

GDC Data Portal: Genomic Data Commons Data Portal

GO: Gene Ontology

NCI: National Cancer Institute d'Estats Units

NHGRI: National Human Genome Research Institute d'Estats Units

TCGA: The Cancer Genome Atlas

TCGAbiolinks: Paquet de R/Bioconductor per a l'anàlisi integratiu de dades del GDC Data Portal

Termes GO: BP, processos biològics; CC, components cel·lulars; MP, funcions moleculars

6. Bibliografia

1. Thompson MJ, Rubbi L, Dawson DW, Donahue TR, Pellegrini M. Pancreatic cancer patient survival correlates with DNA methylation of pancreas development genes. PLoS One. 2015; 10:e0128814.

2. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 2012; 13:484-492.

3. Hughes LAE, Melotte V, de Schrijver J, de Maat M, Smit VTHBM, Bovée JVMG, et al. The CpG island methylator phenotype: what's in a name? Cancer Res. 2013;73:5858–68. doi: 10.1158/0008-5472.CAN-12-4306.

4. Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa J-PJ. CpG island methylator phenotype in colorectal cancer. Proc Natl Acad Sci. 1999;96:8681–6. doi: 10.1073/pnas.96.15.8681.

5. Ueki T, et al. Hypermethylation of multiple genes in pancreatic adenocarcinoma. Cancer Res. 2000;60:1835–1839.

6. Tomczak K, Czerwinska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol (Pozn). 2015; 19:A68-77.

7. Cancer Genome Atlas Research N, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet. 2013; 45:1113-1120.

8. Antonio Colaprico, Tiago Chedraoui Silva, Catharina Olsen, Luciano Garofano, Claudia Cava, Davide Garolini, Thais Sabedot, Tathiane Malta, Stefano M. Pagnotta, Isabella Castiglioni,Michele Ceccarelli, Gianluca Bontempi Houtan Noushmehr. TCGAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data Nucleic Acids Research (05 May 2016) 44 (8): e71. (doi:10.1093/nar/gkv1507).

9. R: Development core team (2018). R: A language and enviroment for

statistical computing. R foundation for statistical computing. Vienna, Austria. http://www.r-project.org/

10. Orchestrating high-throughput genomic analysis with Bioconductor. W. Huber, V.J. Carey, R. Gentleman, ..., M. Morgan. Nature Methods, 2015:12, 115.

11. Gene Ontology Consortium. http://www.geneontology.org/

12. Sánchez-Vega F, Gotea V, Margolin G, et al. Pan-cancer stratification of solid human epithelial tumors and cancer cell lines reveals commonalities and tissue-specific features of the CpG island methylator phenotype. Epigenetics Chromatin. 2015. April;8:14.

13. Goel MK, Khanna P, Kishore J. Understanding survival analysis: Kaplan-meier estimate. Int J Ayurveda Res. 2010;1(4):274. doi: 10.4103/0974-7788.76794.

14. TCGA-Network Integrated genomic characterization of pancreatic ductal adenocarcinoma. Cancer Cell. 2017;32(2):185–203 e113. doi: 10.1016/j.ccell.2017.07.007.

15. Delpu Y, Hanoun N, Lulka H, Sicard F, Selves J, Buscail L, et al. Genetic and epigenetic alterations in pancreatic carcinogenesis. Curr Genomics.2011; 12, 15–24. 10.2174/138920211794520132.

16. Lomberk GA, Iovanna J, Urrutia R. The promise of epigenomic therapeutics in pancreatic cancer. Epigenomics. 2016;8:831–842. doi: 10.2217/epi-2015-0016.

17. Lomberk G, et al. Distinct epigenetic landscapes underlie the pathobiology of pancreatic cancer subtypes. Nat. Commun. 2018;9:1978. doi: 10.1038/s41467-018-04383-6.

7. Annexos

- Llistat de gens hipermetilats en les mostres CIMPH.