

Estudi de la inhibició de ^{V600E}BRAF mitjançant el fàrmac vemurafenib en melanomes

Francesc Turias Fraiso

Màster en Bioinformàtica i Bioestadística
Biologia molecular i estructural

Ivette Olivares Castiñeira

Carles Ventura Royo

24/05/2017



Aquesta obra està subjecta a una llicència de [Reconeixement-NoComercial-SenseObraDerivada 3.0 Espanya de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

FITXA DEL TREBALL FINAL

Títol del treball:	Estudi de la inhibició de ^{V600E} BRAF mitjançant el fàrmac vemurafenib en melanomes
Nom de l'autor:	Francesc Turias Fraisino
Nom del consultor/a:	Ivette Olivares Castiñeira
Nom del PRA:	Carles Ventura Royo
Data de lliurament (mm/aaaa):	05/2017
Titulació o programa:	Màster en Bioinformàtica i Bioestadística
Àrea del Treball Final:	Cerca de mutacions/targets a la via MAPK implicada en varis tipus de càncers
Idioma del treball:	Català
Paraules clau	B-Raf, vemurafenib, melanomes
Resum del Treball (màxim 250 paraules):	
<p>La quinasa B-Raf forma part de la via de transducció del senyal MAPK/ERK, la qual regula la divisió i proliferació cel·lular. Mutacions en aquest proteïna són presents en una gran diversitat de tumors i càncers, especialment destacables en els melanomes, on s'ha trobat mutada al voltant del 70% dels casos.</p> <p>En aquest projecte es pretén avaluar l'efecte de la inhibició de ^{V600E}BRAF –la mutació més freqüent de B-Raf detectada en melanomes- utilitzant el fàrmac vemurafenib, el qual és específic per a aquesta mutació. Per fer-ho s'han dut a terme dos anàlisis estadístics utilitzant el llenguatge de programació R, un a nivell d'expressió gènica i l'altre a nivell poblacional. Utilitzant cèl·lules de melanoma que presenten la mutació V600E, l'objectiu en el primer cas és comparar l'expressió genètica de cèl·lules tractades amb vemurafenib amb la de cèl·lules control. Partint de dades d'expressió de sondes de microarray, aquestes s'han tractat fins obtenir uns resultats interpretables que permeten determinar quins gens i quins processos biològics veuen afectada la seva expressió per l'efecte del vemurafenib. L'objectiu en l'estudi poblacional és comparar la supervivència d'individus tractats amb vemurafenib amb la d'individus tractats amb dacarbazina, avaluant si el vemurafenib suposa una millora en el temps de supervivència.</p> <p>En ambdós casos els resultats obtinguts apunten en la mateixa direcció. L'estudi transcripcional indica una disminució de l'expressió en aquells gens relacionats amb la divisió i la proliferació cel·lular, mentre que l'anàlisi poblacional mostra una major supervivència en els individus tractats amb vemurafenib.</p>	

Abstract (in English, 250 words or less):

B-Raf kinase is part of the signal transduction pathway MAPK/ERK, which regulates cell division and proliferation. Mutations in this protein are present in a wide variety of tumors and cancers, especially in melanomas, which has been found mutated in about 70% of cases.

This project aims to assess the effect of ^{V600E}BRAF inhibition -the most frequent B-Raf mutation detected in melanomas- using the vemurafenib drug, which is specific for this mutation. To do this, we did two statistical analysis using the R programming language, one at genetic expression level and other at the population level. On one hand, transcriptional analysis started with data from microarray probes, which used cells that had the V600E mutation to compare the molecular expression of the ones treated with vemurafenib with the expression of control cells. In this project we have treated the expression data to obtain interpretable results, aiming to determine which genes and biological processes have increased or decreased its expression by the effect of vemurafenib. On the other hand, in the population study we used the survival times of individuals who had the V600E mutation to compare the treated ones with vemurafenib with the individuals treated with dacarbazine, which is unspecific.

In both cases the results point in the same direction. The transcriptional study indicates a decrease in expression of those genes related to cell division and proliferation, while the population analysis indicates greater survival for those individuals treated with vemurafenib.

Índex

1. Introducció.....	5
1.1. Context i justificació del treball.....	5
1.2. Objectius del treball	6
1.3. Enfocament i mètode seguit	6
1.4. Planificació del Treball.....	7
1.4.1. Recursos necessaris	7
1.4.2. Tasques	7
1.4.3. Planificació de les tasques	8
1.5 Breu sumari de productes obtinguts	9
1.6 Breu descripció dels altres capítols de la memòria	9
2. La via MAPK/ERK i la família Raf	11
2.1. Introducció a la via MAPK/ERK i a la transducció del senyal.....	11
2.2. Funció i proteïnes que intervenen en la via MAPK/ERK.....	11
2.3. Com es produeix la cascada de senyals de la via MAPK/ERK	12
2.4. La família de quinases Raf	13
3. La quinasa B-Raf i el fàrmac vemurafenib	15
3.1. Introducció a la quinasa B-Raf.....	15
3.2. Mecanisme d'activació.....	16
3.3. B-Raf i la seva relació amb el càncer	17
3.3.2. La mutació V600E	18
3.4. El fàrmac vemurafenib	19
4. Anàlisi de ^{V600E} BRAF com a diana terapèutica utilitzant el fàrmac vemurafenib en melanomes	22
4.1. Estudi a nivell transcripcional de la inhibició de ^{V600E} BRAF en els melanomes ...	22
4.1.1. Presentació de l'estudi	22
4.1.2. Desenvolupament i resultats de l'estudi	23
4.1.2.1. Lectura de les dades	23
4.1.2.2. Exploració de les dades en cru	25
4.1.2.3. Normalització de les dades.....	31
4.1.2.4. Selecció de gens expressats diferencialment	33
4.1.2.5. Anotació de resultats.....	37
4.1.2.6. Anàlisi de significació biològica	38
4.2. Estudi a nivell poblacional de la inhibició de ^{V600E} BRAF en els melanomes	41
4.2.1. Presentació de l'estudi	41
4.2.2. Desenvolupament i resultats de l'estudi	42
4.2.2.1. Creació de les dades per l'estudi	43
4.2.2.2. Càlcul de l'estimador de Kaplan-Meier	46
4.2.2.3. Estimació de les corbes de supervivència i càlcul del <i>Log-rank test</i>	48
5. Conclusions	50
5.2. Conclusions del treball	50

5.2. Assoliment dels objectius plantejats.....	51
5.3. Seguiment de la planificació i metodologia	51
5.4. Línies de treball futur	52
6. Referències	53
7. Annex	58

Llista de figures

Capítol 2. La via MAPK/ERK i la família Raf

Figura 2.1. Via MAPK/ERK de transducció del senyal (imatge de Kim HJ i Bar-Sagi D, 2004)^[36].

Capítol 3. La quinasa B-Raf i el fàrmac vemurafenib

Figura 3.1. Estructura terciària de B-Raf. Imatge obtinguda del servidor web PDB. PDB ID: 1UWH^[44].

Figura 3.2. Estructura secundària de l'àcid glutàmic (a) i de la serina fosforilada (b). Les estructures s'han creat mitjançant el navegador web ChEMBL^[45].

Figura 3.3. Estructura secundària del fàrmac vemurafenib. L'estructura s'ha creat mitjançant el navegador web ChEMBL^[45].

Figura 3.4. Estructura terciària del centre actiu de B-Raf (en forma de monòmer) amb el fàrmac vemurafenib unit. Imatge obtinguda del servidor web PDB. PDB ID: 4RZV^[39].

Capítol 4. Anàlisi de ^{V600E}BRAF com a diana terapèutica utilitzant el fàrmac vemurafenib en melanomes

Figura 4.1. Histograma de la distribució de les expressions de les sis mostres d'estudi.

Figura 4.2. Boxplot de la distribució de les expressions de les sis mostres d'estudi.

Figura 4.3. Agrupament jeràrquic de les sis mostres d'estudi.

Figura 4.4. Gràfic d'anàlisi de components principals de les sis mostres d'estudi.

Figura 4.5. Gràfic MAplot de les sis mostres d'estudi.

Figura 4.6. Gràfic d'expressió relativa (RLE) de les sis mostres d'estudi.

Figura 4.7. Gràfic d'error no estandarditzat i normalitzat (NUSE) de les sis mostres.

Figura 4.8. Boxplot de la distribució de les expressions de les sis mostres normalitzades.

Figura 4.9. Gràfic MAplot de les sis mostres normalitzades.

Figura 4.10. Gràfic d'anàlisi de components principals de les sis mostres normalitzades.

Figura 4.11. Volcano plot dels gens expressats. Els valors superior a (-1,1) a l'eix de les X (Log Fold Change) i superiors a 0 a l'eix de les Y (Log Odds) són els que es troben diferencialment expressats.

Figura 4.12. Heatmap dels gens diferencialment expressats entre les mostres tractades amb vemurafenib i les mostres control.

Figura 4.13. Gràfic de les corbes de supervivència pels individus tractats amb vemurafenib (línia negra) i dacarbazine (línia grisa) obtingut de l'estudi dut a terme per Chapman et al^[41]. Les línies verticals de les corbes representen els individus censurats durant el temps d'estudi. A l'eix de les X es mostra el temps de supervivència (mesos) i a l'eix de les Y el percentatge de supervivència.

Figura 4.14. Gràfic d'estimació de les corbes de supervivència pels individus tractats amb vemurafenib (línia blava) i dacarbazine (línia grana). Les línies verticals de les corbes representen els individus censurats durant el temps d'estudi. A l'eix de les X es mostra el temps de supervivència (mesos) i a l'eix de les Y el percentatge de supervivència.

1. Introducció

1.1. Context i justificació del treball

El càncer és una malaltia amb una incidència notable i sovint mortal en la població. Múltiples mutacions i desregulacions dins les cèl·lules poden desembocar en l'aparició de tumors i càncers, motiu pel qual avui dia és una tasca inabastable la creació d'una cura general i efectiva pel càncer. Tot i això, cada vegada s'estan descobrint i creant més fàrmacs específics per combatre els càncers, amb els quals, en general, s'obtenen uns resultats més satisfactoris^[30].

En aquest projecte es pretén estudiar l'efecte de la inhibició de $V600E$ BRAF, la mutació més comuna de la quinasa B-Raf, en melanomes metastàtics mitjançant la utilització del fàrmac vemurafenib, el qual és específic per a aquesta mutació. El projecte s'ha centrat en els melanomes ja que B-Raf es troba mutada en el 60-70% dels casos en aquest tipus de càncer^{[12][13]}, fet que la converteix en una possible diana a tenir molt en compte. A més, els melanomes metastàtics són un dels tipus més perillosos de càncer, amb una supervivència que oscil·la entre 8 i 18 mesos en pacients acabats de diagnosticar^{[30][41]}.

L'efectivitat del tractament s'ha valorat a nivell transcripcional i a nivell poblacional, tot i que primer s'ha aprofundit en el coneixement tant de B-Raf (i la mutació $V600E$ BRAF) com del fàrmac vemurafenib. Per una banda, a l'anàlisi transcripcional s'ha utilitzat la línia cel·lular A375 (cèl·lules de melanoma que presenten la mutació $V600E$ BRAF) i s'ha comparat l'expressió gènica de cèl·lules tractades amb vemurafenib amb l'expressió de cèl·lules control. Partint de les dades d'expressió de les sondes de microarray, s'ha dut a terme tot el procés d'anàlisi per obtenir uns resultats interpretables i poder determinar quins gens i processos cel·lulars es veuen inhibits i sobreexpressats per l'efecte del vemurafenib. Per altra banda, en l'anàlisi poblacional s'ha realitzat un anàlisi de supervivència comparant individus tractats amb vemurafenib amb individus tractats amb dacarbazine (medicament inespecífic), els quals presenten càncer de melanoma i la mutació $V600E$. Amb això es pretén avaluar si la supervivència dels individus és major en aquells que segueixen el tractament amb vemurafenib.

1.2. Objectius del treball

1. Profunditzar en l'estudi de la quinasa B-Raf i conèixer el fàrmac vemurafenib, introduint prèviament la via MAPK/ERK i la família de quinases Raf.
 - a. Estudiar la via MAPK/ERK: funció, proteïnes que hi intervenen i explicació de com es produeix la cascada de senyals.
 - b. Conèixer la família de quinases Raf (A-Raf, B-Raf i Raf-1), la qual actua en la via MAPK/ERK.
 - c. Introduir i analitzar computacionalment el gen BRAF i la quinasa B-Raf a nivell de seqüència i estructura, localització en el genoma humà i número d'exons.
 - d. Entendre el mecanisme d'activació de B-Raf.
 - e. Estudiar la relació de B-Raf amb els càncers i conèixer la mutació ^{V600E}BRAF.
 - f. Introduir el fàrmac vemurafenib, el qual té la mutació V600E com a diana terapèutica.
2. Estudiar l'efecte de la inhibició de ^{V600E}BRAF mitjançant l'ús del fàrmac vemurafenib.
 - a. Comparar l'expressió diferencial de gens en cèl·lules de la línia A375 tractades amb vemurafenib amb l'expressió de cèl·lules A375 sense tractament, avaluant aquells gens i processos cel·lulars que es troben *up* o *down*-regulats per l'ús del fàrmac.
 - b. Comparar les corbes de supervivència d'individus tractats amb vemurafenib amb la d'individus tractats amb dacarbazine, avaluant si el tractament amb vemurafenib augmenta la supervivència dels individus amb melanoma.

1.3. Enfocament i mètode seguit

L'apartat d'anàlisi de dades (corresponent a l'objectiu número 2) és on poden existir diverses estratègies a seguir. Per començar, en ambdós casos (tant en l'anàlisi transcripcional com en el poblacional) s'ha utilitzat el llenguatge de programació

estadístic R (RStudio) per dur-los a terme, ja que ofereix un gran ventall de paquets que permeten realitzar tasques molt diverses i elaborades.

Per dur a terme l'anàlisi de dades provinents de microarray s'ha utilitzat el paquet *oligo*^[40]. Tot i que existeixen diversos paquets que també permeten dur a terme anàlisi de sondes de microarray, com *edgeR* o *DESeq*, s'ha optat pel paquet *oligo* perquè és més adequat per treballar amb dades d'oligonucleòtids, com és el cas. També s'han utilitzat altres paquets necessaris per completar l'estudi i que es comentaran més endavant. Per l'anàlisi de supervivència només ha estat necessari utilitzar el paquet *survival*^[53], el qual ofereix les funcions i eines suficients per realitzar un anàlisi complet.

1.4. Planificació del Treball

1.4.1. Recursos necessaris

Gran part del projecte s'ha dut a terme mitjançant el llenguatge de programació estadístic R (RStudio), el qual ha servit per realitzar tant l'anàlisi de dades d'expressió provinents de microarrays com l'anàlisi de supervivència.

La major part d'informació bibliogràfica sobre la via MAPK/ERK, B-Raf i vemurafenib ha estat obtinguda d'articles científics, mentre que per la part computacional de B-Raf i vemurafenib s'han usat els servidors genòmics UCSC (<https://genome.ucsc.edu>)^[32], UniProt (www.uniprot.org)^[59], ChEMBL (www.ebi.ac.uk/chembl)^[45] i Protein Data Bank (www.rcsb.org)^[58].

1.4.2. Tasques

Tasca 1 (corresponent a l'objectiu número 1):

- Tasca 1.1. Cerca bibliogràfica sobre la via MAPK/ERK en articles científics. Duració: 5 dies (del 16/03 al 20/03).
- Tasca 1.2. Cerca bibliogràfica sobre la família RAF en articles científics. Duració: 4 dies (del 21/03 al 24/03).
- Tasca 1.3. Cerca bibliogràfica en articles científics totalment centrada en la quinasa B-Raf i utilització de servidors genòmics i proteics per a realitzar un anàlisi computacional del gen i la proteïna. Duració: 8 dies (del 25/03 al 01/04).
- Tasca 1.4. Cerca bibliogràfica i utilització de servidors genòmics centrada en la mutació V600E. Duració: 3 dies (del 03/04 al 05/04).

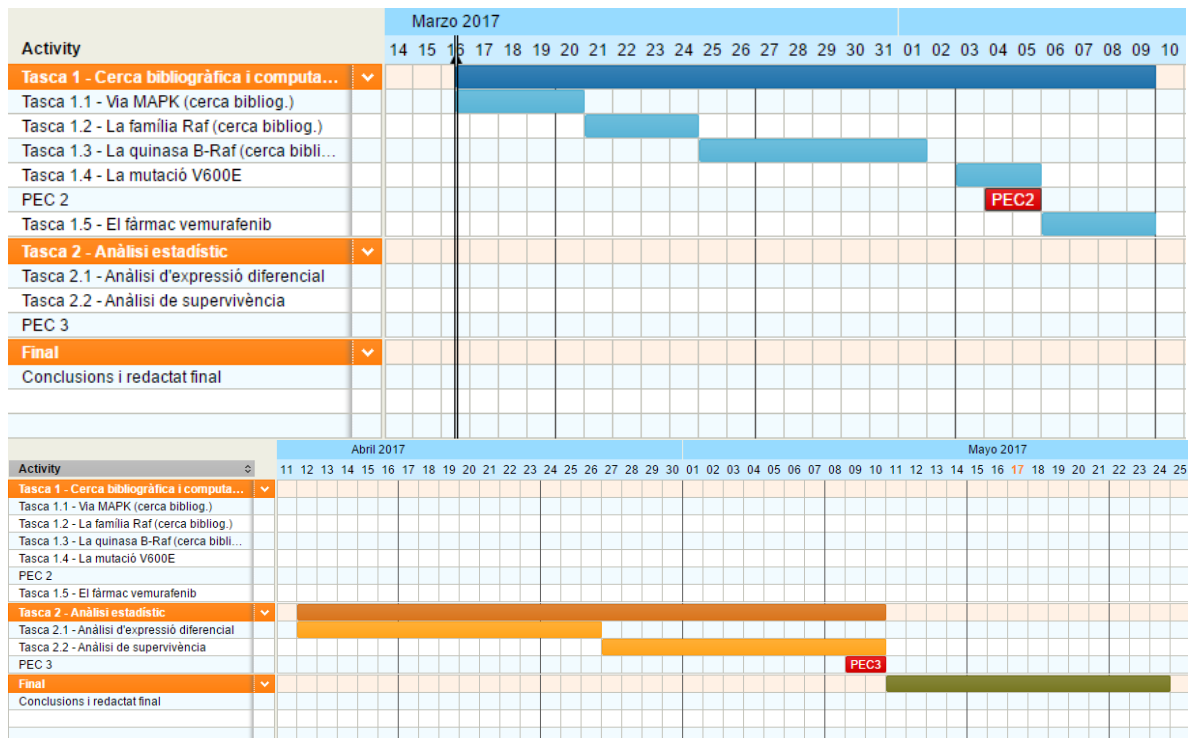
- Tasca 1.5. Cerca bibliogràfica i en servidors genòmics del fàrmac vemurafenib. Duració: 4 dies (del 06/04 al 09/04).

Tasca 2 (corresponent a l'objectiu número 2):

- Tasca 2.1. Utilització del llenguatge de programació R per a realitzar l'anàlisi d'expressió diferencial de les dades obtingudes mitjançant microarray. El procés consta, de manera general, de: anàlisi de les dades en cru, normalització de les dades, selecció i anotació de gens diferencialment expressats i anàlisi de significació biològica. Duració: 15 dies (del 12/04 al 26/04).
- Tasca 2.2. Utilització del llenguatge de programació R per a realitzar l'anàlisi de supervivència: estimació d'individus fallats i censurats, estimació de les corbes de supervivència dels dos grups i contrast d'hipòtesis. Duració: 14 dies (del 27/04 al 10/05).

1.4.3. Planificació de les tasques

La planificació temporal de les tasques es mostra a continuació:



1.5 Breu sumari de productes obtinguts

La llista de resultats i productes obtinguts és la següent:

- Memòria.
- Els productes següents:
 - **AnalisiMicroarray**: script d'R amb els passos realitzats en l'anàlisi d'expressió diferencial de gens.
 - **AnalisiSupervivencia**: script d'R amb els passos realitzats en l'anàlisi de supervivència.
 - **AnotacioGensUp** i **AnotacioGensDown**: arxius html recollint els gens (i les seves característiques) que es troben diferencialment expressats, essent el primer arxiu pels gens *up*-regulats i el segon pels gens *down*-regulats.
 - **GOup** i **GOdown**: arxius html amb la llista de termes GO més freqüents en els gens diferencialment expressats, essent el primer arxiu pels termes *up*-regulats i el segon pels termes *down*-regulats.
 - **KEGGup** i **KEGGdown**: arxius html amb la llista de termes KEGG més freqüents en els gens diferencialment expressats, essent el primer arxiu pels termes *up*-regulats i el segon pels termes *down*-regulats.
- Presentació virtual.

1.6 Breu descripció dels altres capítols de la memòria

Capítol 2: La via MAPK/ERK i la família Raf. En aquest capítol s'introdueix la via MAPK/ERK, les proteïnes que hi intervenen i s'explica com es produeix la cascada de transducció del senyal en aquesta via. També s'introdueix la família de quinases Raf, les quals participen en la via MAPK/ERK. Tot aquest capítol posa en context la quinasa B-Raf, protagonista del projecte, i ajuda a entendre per què mutacions en aquesta proteïna poden desembocar en una proliferació incontrolada de les cèl·lules.

Capítol 3: La quinasa B-Raf i el fàrmac vemurafenib. En aquest capítol es du a terme un estudi de la quinasa B-Raf, començant per un breu anàlisi computacional tant del gen com de la proteïna, seguint amb el seu mecanisme d'activació i acabant explicant

la seva relació amb el càncer, introduint així la mutació més freqüent de la proteïna, ^{V600E}BRAF. Finalment també s'analitza el fàrmac vemurafenib, el qual és selectiu per aquesta mutació i és el medicament emprat en els anàlisis posteriors.

Capítol 4: Anàlisi de ^{V600E}BRAF com a diana terapèutica utilitzant el fàrmac vemurafenib en melanomes. En aquest capítol es duen a terme dos anàlisis estadístics mitjançant R, un a nivell transcripcional i l'altre a nivell poblacional, per tal d'avaluar l'efecte de la inhibició de BRAF-V600E utilitzant el fàrmac vemurafenib. Per una banda, partint de les dades d'expressió de cèl·lules A375 tractades i no tractades amb vemurafenib, es realitza tot el procediment d'anàlisi de sondes de microarray fins obtenir una col·lecció de gens que es troben diferencialment expressats entre les dues condicions, permetent determinar l'efecte del vemurafenib a nivell genètic. Per altra banda, es compara el temps de supervivència d'individus tractats amb vemurafenib amb el d'individus tractats amb dacarbazine, tots ells presentant la mutació V600E, per tal d'avaluar si el fàrmac vemurafenib representa o no una millora en comparació amb els fàrmacs tradicionals.

2. La via MAPK/ERK i la família Raf

2.1. Introducció a la via MAPK/ERK i a la transducció del senyal

La via MAPK/ERK, també anomenada via Ras-Raf-MEK-ERK^[1], és una de les múltiples vies de transducció del senyal que existeixen dins les cèl·lules. Totes les funcions que duen a terme les cèl·lules estan regulades per senyals, ja siguin extracel·lulars o intracel·lulars, i en el moment en què una cèl·lula en deixa de rebre entra en el procés de mort programada anomenat apoptosi. Perquè aquests senyals tinguin efecte en la cèl·lula és vital que la informació arribi al nucli, fet que s'aconsegueix mitjançant les vies de senyalització cel·lular.

El procés s'inicia quan un primer missatger, que pot variar des d'una hormona fins a un gas simple com el CO, interacciona amb un receptor de membrana. Aquesta interacció posa en marxa la transducció del senyal, que consisteix en una cadena de proteïnes quinases que es van fosforilant (i desfosforilant) seqüencialment, fins que el senyal arriba al nucli i produeix un canvi fisiològic en la cèl·lula.

La via MAPK/ERK forma part del conjunt de vies de senyalització que tenen en comú l'activació d'una proteïna MAPK (de l'anglès *mitogen-activated protein kinase*)^[2]. Les proteïnes MAPK més estudiades i conegudes són JNK, p38 i ERK, la qual té relació amb aquest estudi.

2.2. Funció i proteïnes que intervenen en la via MAPK/ERK

Com s'ha explicat, les funcions que es produeixen dins les cèl·lules estan regulades per senyals. Aquests varien depenent de la via de transducció que s'activi, però les vies no estan lligades a una sola funció. Una via pot regular diverses funcions de la mateixa manera que una sola funció pot estar regulada per diverses vies. Així, la via MAPK/ERK té un paper essencial en la regulació de processos com la divisió i la proliferació cel·lular, i també s'ha comprovat que intervé en la regulació d'altres funcions com l'apoptosi i la migració cel·lular^{[3][9]}. És lògic pensar que, en ser una via important per a la proliferació cel·lular, mutacions en alguna de les proteïnes implicades pugui desembocar en una proliferació incontrolada de la cèl·lula.

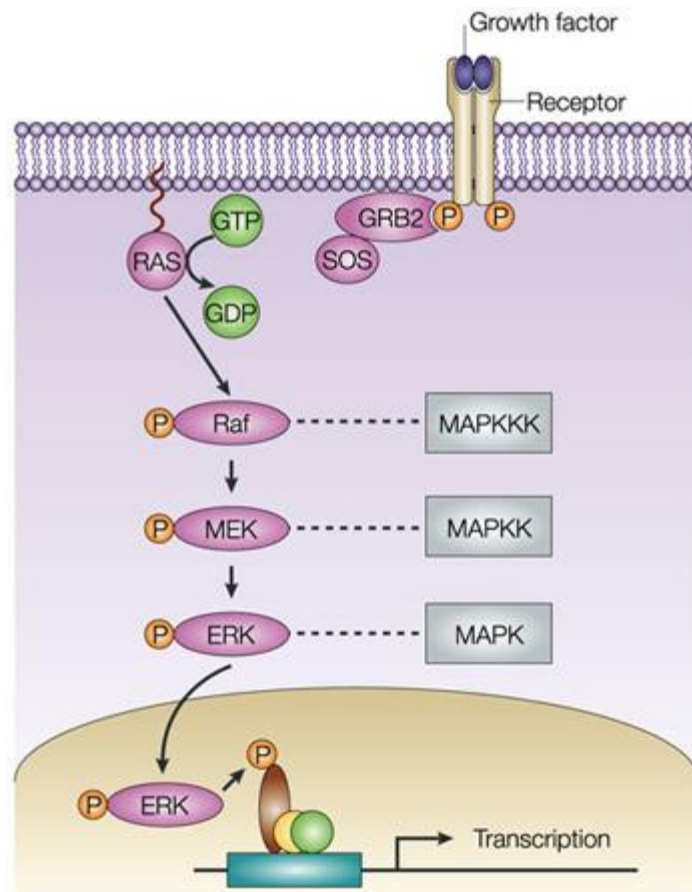
De manera general, les vies MAPK involucren una proteïna G, una MAPKKK (*mitogen-activated protein kinase kinase kinase*), una MAPKK i una MAPK^[4]. En la via MAPK/ERK, aquestes quinases són les mateixes proteïnes que li donen nom a la via: Ras, Raf, MEK (MEK1/2) i ERK (ERK1/2), respectivament. Així, Ras actua com a proteïna G, Raf com a MAPKKK, MEK actua com a MAPKK i, finalment, ERK com a MAPK. Tot i això, no s'ha de veure una via de senyalització com un circuit tancat. De la mateixa manera que un senyal pot activar més d'una via i viceversa, hi ha altres proteïnes que interaccionen amb les quinases de la cadena ajudant en l'activació i la regulació^[7]. Tot i que més endavant es veuran exemples d'aquest fet, per explicar el funcionament de la via no es tindran en compte.

2.3. Com es produeix la cascada de senyals de la via MAPK/ERK

La cascada de senyals és un procés que implica tota una cadena de proteïnes quinases, les quals es van activant seqüencialment fins fer arribar el senyal al nucli. En estat de repòs les quinases es troben desfosforilades, i la seva activació es produeix mitjançant la fosforilació d'un o més residus de la cadena polipeptídica, habitualment serines, tirosines o treonines^[6]. En activar-se fosforilen la següent quinasa de la cadena, tornant així a l'estat inactiu i permetent que segueixi la cascada de transducció.

En el cas de la via MAPK/ERK, aquesta s'inicia quan un lligand s'uneix al receptor de membrana. Això provoca una dimerització del receptor i la fosforilació d'una tirosina que hi està unida. Dues proteïnes lligades al receptor, GRB2 i SOS, reconeixen aquesta tirosina fosforilada i permeten l'activació de Ras, la qual allibera una GDP i capta una GTP, passant a un estat actiu. Mentre tingui una GTP unida podrà activar Raf, la qual fosforilarà MEK i, finalment, MEK activarà ERK, que entrarà dins del nucli i fosforilarà un factor de transcripció^[1]. La seqüència de la via de transducció MAPK/ERK es pot observar a la Figura 2.1.

Com s'ha explicat, un cop es produeix la fosforilació d'una quinasa, la quinasa anterior es desfosforila i torna a l'estat inactiu. Aquest mètode de regulació impedeix que hi hagi una sobreactivació de la via i una proliferació descontrolada de les cèl·lules.



Nature Reviews | Molecular Cell Biology

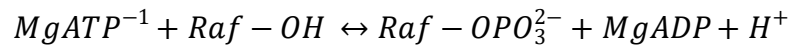
Figura 2.1. Via MAPK/ERK de transducció del senyal (imatge de Kim HJ i Bar-Sagi D, 2004)^[36].

2.4. La família de quinases Raf

Un dels enzims que intervé en la cascada de senyals MAPK/ERK és la proteïna Raf. Raf, de l'anglès *Rapidly Accelerated Fibrosarcoma*^[9], és una família de serina/treonina quinases que, en mamífers, consta de tres isoformes: A-Raf, B-Raf i Raf-1 (o C-Raf)^{[5][7]}. Totes tres tenen una alta homologia, i la principal diferència entre elles rau en la força d'activació de MEK, essent B-Raf la que té una força d'activació major i A-Raf la menor^[6].

La família de quinases Raf presenta tres dominis proteics: CR1, CR2 i CR3 (de l'anglès *Conserved Regions*). El domini CR1 és el que conté el domini d'unió amb Ras, mentre que CR3 conté l'anomenat *kinase activation loop*^[11], lloc on es produeix l'activitat quinasa. La fosforilació dels residus situats en aquest *loop* del domini CR3 és vital per a l'activació de la quinasa. Finalment, el domini CR2 regula tant la unió amb Ras com

l'activació de Raf a través de residus inhibitoris mitjançant fosforilació. Els dominis CR1 i CR2 se situen al N-terminal, mentre que el CR3 es troba al C-terminal de la proteïna^{[5][7][11]}. La reacció que catalitza Raf per a la seva activació és la següent^[5]:



Tot i que la primera quinasa Raf que es va descobrir fou Raf-1^[7], la importància d'aquesta família va créixer notablement amb el descobriment i l'observació de B-Raf mutades en diversos tipus de tumors i càncers. La influència de les mutacions de B-Raf és notable i especialment destacable en els melanomes, on se sol trobar mutada entre el 50 i el 70% dels casos^{[12][13][14]}. Aquests dos fets converteixen B-Raf en una possible diana terapèutica a tenir en compte i en un dels objectius principals d'aquest projecte.

3. La quinasa B-Raf i el fàrmac vemurafenib

La quinasa B-Raf, juntament amb el fàrmac vemurafenib, són els dos elements principals en què se centra aquest estudi. L'objectiu final és avaluar si la inhibició de la mutació ^{V600E}BRAF mitjançant el fàrmac vemurafenib representa una benefici i una millora en el tractament dels melanomes. No obstant això, abans de començar amb els anàlisis pot ser interessant conèixer més profundament els dos protagonistes, especialment els aspectes que contribueixen a entendre la implicació de B-Raf en els càncers i posar en context els anàlisis posteriors, començant amb una breu introducció a la quinasa B-Raf.

3.1. Introducció a la quinasa B-Raf

La serina/treonina quinasa B-Raf forma part del conjunt de proteïnes que actuen a la via de transducció MAPK/ERK. És codificada per l'oncogen BRAF i, com s'ha explicat, està involucrada en el creixement i desenvolupament cel·lular. La quinasa B-Raf és activada per Ras mitjançant fosforilació, i en el seu estat actiu forma un dímer a través d'enllaços d'hidrogen en el *kinase activation loop*^[37]. Això permet activar la quinasa MEK i seguir amb la transducció del senyal.

El gen BRAF es troba en el braç *q* del cromosoma 7 del genoma humà, consta de 18 exons (les parts que, posteriorment, es traduiran a proteïna) i se situa entre les posicions 140,734,479 i 140,924,764 de la cadena nucleotídica^{[32][43]}. A l'Annex s'adjunta la seqüència nucleotídica de la regió codificant de BRAF. El gen BRAF codifica per la quinasa B-Raf, formada per una cada proteica de 766 aminoàcids^{[32][43]}, essent la seqüència sencera la següent:

```
>uc003vwc.5 (BRAF) length=766
MAALSGGGGGAEPGQALFNGDMEPEAGAGAGAAASSAADPAIPEEVNLIKQMIKLTQEH
IEALLDKFGGEHNPPSIYLEAYEYTSKLDALQQREQLLES LGNGTDFSVSSASMDTV
TSSSSSSLSVLPSSLVFNQPTDVARSNPKSPQKPIVRVFLPNKQRTVVPARCGVTVRDS
LKKALMMRGLIPECCAVYRIQDGEKKPIGWDTDISWL TGEELHVEVLENVPL TTHNFVRK
TFFT LAFCDFCRKL LFGFRQCQTCGYKFHQRCSTEVP LMCVNYDQLDLLFVSKFFEHP I
PQEEASLAETAL TSGSSPSAPASDSIGPQILTSPSPSKSIPQPFPADEDRNQFGQR
DRSSAPNVHINTIEPNIDDLIRDQGFGRDGGSTTGLSATPPASLPGLTINVKALQKSP
GPQREKSSSSSEDRNRMKTLGRRDSSDDWEIPDQGITVQQRIGSGSFGTVYKKGKWHGDV
AVKMLNVTAPTQQQLQAFKNEVGLRKRHVNI LLFMGYSTKPLAIVTQWCEGSSLYHH
LHIIETKFEMIKLIDIAEQTAQMDYLHAKSIIHRDLKSNIFLHEDLTVKIGDFGLATV
KSRWSSGSHQFEQLSGSILWMAPEVIRMQDKNPYSFQSDVYAFGIVLYELMTGQLPYSNIN
NRDQIIFMVGRGYLSPDL SKVRSNCPKAMKRLMAECLKKRDERPLFPQILASIELLARS
LPKIHRSAEPLNLRAGFQTEDFS LYACASPKTPIQAGGYGAFVH
```

Finalment, l'estructura terciària de la quinasa B-Raf es mostra a la Figura 3.1, la qual està formada per 19 hèlix alfa i 16 làmines beta.

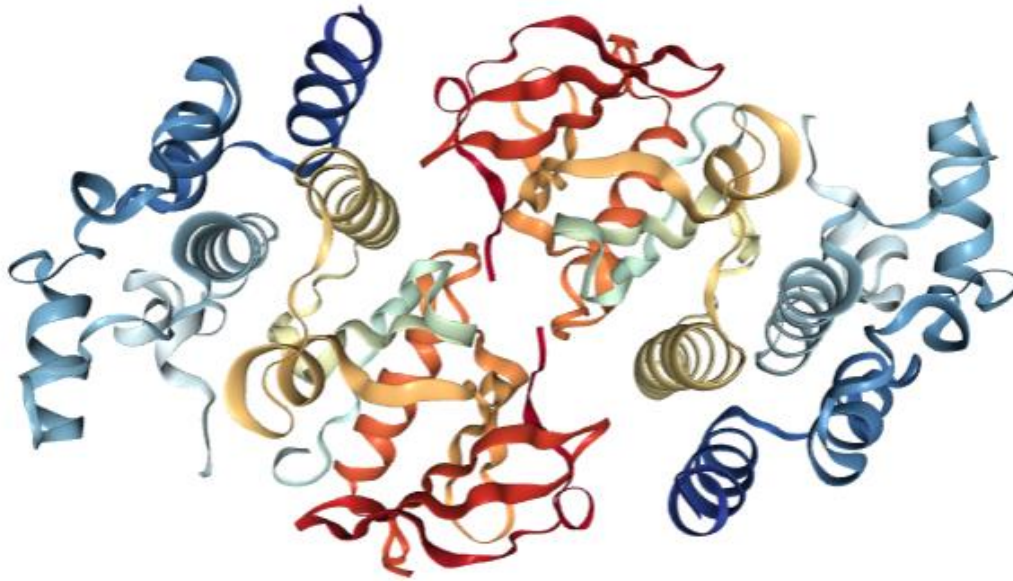


Figura 3.1. Estructura terciària de B-Raf. Imatge obtinguda del servidor web PDB. PDB ID: 1UWH^[44].

3.2. Mecanisme d'activació

Un cop introduïda la quinasa B-Raf a nivell de seqüència i estructura, en aquest apartat s'exposarà el seu mecanisme d'activació. Conèixer quins residus són fosforilats per a l'activació de B-Raf i la seva importància en el procés servirà de base per entendre la mutació V600E, que es veurà en el següent apartat.

El mecanisme d'activació, en general, és similar pels tres membres de la família Raf, tot i que difereixen en alguns aspectes. Les quinases Raf, com totes les quinases que participen a la cadena MAPK/ERK, s'activen mitjançant la fosforilació de residus aminoacídics, principalment serines, treonines i tirosines^[6]. En el cas de B-Raf, la seva activació és produïda per la fosforilació dels residus de treonina en la posició 598 (Thr598) i de serina en la posició 601 (Ser601), els quals se situen en el *kinase activation loop* comentat anteriorment^{[6][11]}. Aquests dos residus es troben conservats en els tres membres de la família, tot i que en diferents posicions de la cadena (a C-Raf, per exemple, es localitzen en la posició Thr491 i Ser494)^[6].

Tot i que B-Raf presenta altres llocs de fosforilació, els quals ajuden a regular la seva activitat, la fosforilació dels residus Thr598 i Ser601 és suficient per a l'activació de B-

Raf. Sense la fosforilació d'aquests dos residus no hi activació^[6], fet que els converteix en aminoàcids essencials per entendre el mecanisme d'activació de B-Raf. A més, aquest tret diferencia B-Raf de les altres dues quinases de la família, Raf-1 i A-Raf, ja que aquestes necessiten una doble fosforilació per al seu funcionament: a part de l'activació produïda per Ras en els dos residus mencionats, necessiten ser activades també per la quinasa Src^[6]. El fet de tenir una activació més forta que les altres dues membres de la família Raf és un dels motius pels quals la seva implicació en diversos càncers és major.

3.3. B-Raf i la seva relació amb el càncer

Com s'ha explicat, B-Raf està implicada en una via de senyalització que regula el creixement i la divisió cel·lular, així que és lògic pensar que mutacions en aquesta proteïna poden comportar una divisió incontrolada de les cèl·lules. Òbviament l'aparició d'un càncer no depèn només de la mutació d'un sol gen^[13], però el percentatge de mutació que pateix el gen en alguns tipus de càncer és indicatiu de la seva importància per al desenvolupament de la malaltia, essent B-Raf un d'aquests casos.

La importància que té B-Raf en els càncers es va començar a descobrir a principis dels anys 2000^[7], i des de llavors s'han realitzat nombrosos treballs on s'ha estudiat la seva implicació en diversos tipus de tumors i càncers. Mutacions en la seqüència de BRAF s'han observat en melanomes, càncers papil·lars tiroïdals, càncers colorectals, càncers de pulmó i fetge, tumors de l'estroma gastrointestinal, sarcomes, càncers de mama, gliomes i càncers ovàrics^{[12][13][17][18][19][20][21][22]}. No obstant això, on sens dubte tenen més incidència les mutacions en BRAF són en els melanomes i en els càncers papil·lars tiroïdals, trobant-se mutada al voltant del 60% i 40% dels casos, respectivament^{[13][19]}. A més, B-Raf és la proteïna que es troba mutada més habitualment en els melanomes, indicant la importància d'aquesta per al desenvolupament de la malaltia^[33]. En tots els altres tipus de tumors i càncers esmentats les mutacions a B-Raf se situen al voltant del 10% dels casos, essent inclús inferior al 1% en càncers com el de l'estroma gastrointestinal^[20].

La gran majoria de mutacions que pateix BRAF, independentment del tipus de càncer, es deuen per la substitució d'un nucleòtid per un altre en la cadena de DNA. De totes les mutacions conegudes de B-Raf en càncers, la que sens dubte destaca per sobre les altres és la mutació V600E. Aquesta és la més comuna trobada en els diversos tipus de càncers, i és la mutació que pateixen les cèl·lules i els individus en els anàlisis que es duran a terme posteriorment en aquest projecte.

Com apunt final, les mutacions a BRAF no només poden conduir a l'aparició de càncers. També s'han detectat mutacions en el gen BRAF (així com en altres proteïnes de la via MAPK/ERK) en malalties hereditàries com la síndrome cardiofaciocular (*CFC syndrome*) o la síndrome LEOPARD^{[15][16]}.

3.3.2. La mutació V600E

La mutació V600E, ^{V600E}BRAF o BRAF-V600E, és la mutació més comuna de B-Raf. Es produeix per la substitució d'una timina per una adenina en la posició 1799 (exó 15) de la cadena nucleotídica, que provoca el canvi d'una valina (V) per un àcid glutàmic (E) en el residu 600 de la proteïna resultant^[10].

Aquesta substitució és majoritària en gran part dels càncers mencionats, arribant fins al 90% dels casos en melanomes^{[8][10]}. En càncers papil·lars tiroïdals aquesta proporció és menor, variant del 35% al 70% depenent de l'estudi^{[19][23]}, però en tots ells és la substitució predominant. On sorgeixen més dubtes és amb el càncer de pulmó, ja que s'ha observat que aquesta mutació només representa al voltant d'un 10% de les mutacions de BRAF^[18], tot i que en estudis recents aquesta proporció puja fins el 50%, essent també la substitució predominant^[24].

És destacable, doncs, el pes d'aquesta substitució en la formació de tumors i la posterior aparició de càncers, així com lògic preguntar-se per què concretament aquesta mutació és tan freqüent. La resposta es troba en la unió de dos fets claus: el lloc on es produeix la substitució i l'aminoàcid que substitueix la valina. Com s'ha comentat anteriorment, l'activació de B-Raf es dona per la fosforilació dels residus Thr598 i Ser601, i sense la fosforilació d'aquests dos aminoàcids no hi ha activació de la quinasa. Així, el primer fet clau és que la substitució es produeix just en el *kinase activation loop*, entre els dos residus que es fosforilen per a l'activació^[10]. El segon fet clau és que l'aminoàcid que substitueix la valina és l'àcid glutàmic, el qual actua com a

fosfomimètic^{1[10]}. A la Figura 3.2 es pot observar i comparar l'estructura secundària d'un àcid glutàmic amb la d'una serina fosforilada.

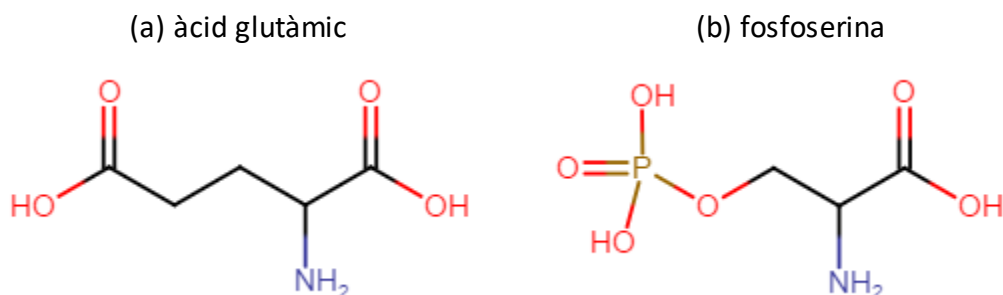


Figura 3.2. Estructura secundària de l'àcid glutàmic (a) i de la serina fosforilada (b). Les estructures s'han creat mitjançant el navegador web ChEMBL^[45].

Per tant, el fet de col·locar un aminoàcid fosfomimètic en el centre actiu de la quinasa provoca que B-Raf estigui constantment activada. Això implica una fosforilació continua de MEK i una activació permanent de la via, conduint a una proliferació descontrolada de la cèl·lula.

La gran presència de la mutació ^{V600E}BRAF en els càncers la converteix en una possible diana terapèutica a tenir molt en compte, i des de fa relativament poc temps han començat a sortir medicaments específics per aquesta mutació, com per exemple el fàrmac vemurafenib.

3.4. El fàrmac vemurafenib

Com s'ha pogut veure fins ara, la mutació BRAF-V600E té una alta incidència entre els melanomes. Tot i així, fins fa relativament poc només existien dos tractaments per a aquesta malaltia: interleuquina-2 i dacarbazine; els quals, a més, eren inespecífics i presentaven una baixa eficàcia^[30]. L'aprovació el 2011 del fàrmac vemurafenib^[30] va suposar l'aparició del primer medicament específic per combatre aquesta mutació i un salt qualitatiu per al tractament de melanomes.

Vemurafenib, també anomenat PLX4032^[25], és una molècula selectiva per a la mutació V600E, tot i que també actua contra la mutació V600K que apareix amb menys

¹ Aminoàcid fosfomimètic: aminoàcid que presenta una estructura terciària molt semblant a un aminoàcid fosforilat.

freqüència en els melanomes (al voltant d'un 10% dels casos^[38]). Bàsicament, el fàrmac vemurafenib actua unint-se al *kinase activation loop* de ^{V600E}BRAF impedit que pugui dur a terme la seva activitat. La seva fórmula molecular és C₂₃H₁₈ClF₂N₃O₃S i l'estructura secundària es presenta a la Figura 3.3.

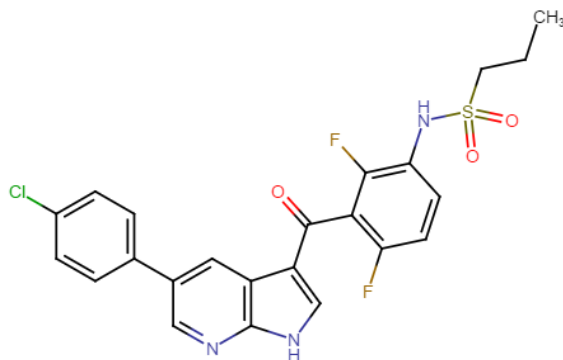


Figura 3.3. Estructura secundària del fàrmac vemurafenib. L'estructura s'ha creat mitjançant el navegador web ChEMBL^[45].

A la Figura 3.4 es presenta l'estructura terciària del centre actiu de B-Raf amb el fàrmac vemurafenib unit.

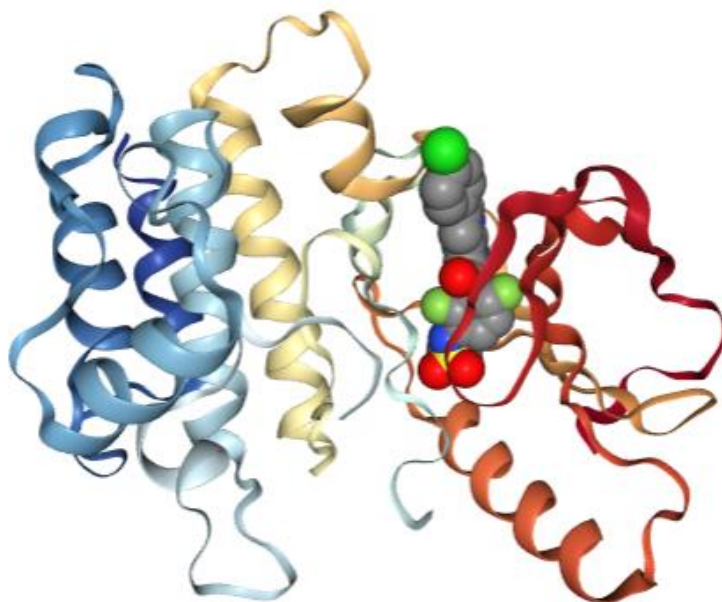


Figura 3.4. Estructura terciària del centre actiu de B-Raf (en forma de monòmer) amb el fàrmac vemurafenib unit. Imatge obtinguda del servidor web PDB. PDB ID: 4RZV^[39].

L'aparició d'un càncer no es deu només a la mutació d'un sol gen. Dins la cèl·lula es produeixen tot un seguit d'alteracions i desregulacions de diversos gens que desemboquen en la malaltia. Tot i això, s'ha demostrat que és possible provocar la mort de cèl·lules canceroses inhibint només una via metabòlica crítica^[8]. Per tant, és moment de comprovar si la inhibició de ^{V600E}BRAF mitjançant el fàrmac vemurafenib té uns efectes beneficiosos per combatre els melanomes.

4. Anàlisi de ^{V600E}BRAF com a diana terapèutica utilitzant el fàrmac vemurafenib en melanomes

Un cop coneguts els dos protagonistes és moment de comprovar si l'ús del fàrmac vemurafenib és efectiu per combatre els melanomes que presenten la mutació ^{V600E}BRAF. Per valorar aquesta eficàcia es duran a terme dos anàlisis, un a nivell transcripcional i l'altre a nivell poblacional, per tal d'avaluar com afecta el fàrmac a l'expressió dels gens i de quina manera repercuteix això en el temps de supervivència de l'individu. En el primer estudi es comprovarà quins gens i funcions cel·lulars es veuen afectades per l'ús del vemurafenib, comparant l'expressió genètica de cèl·lules canceroses sense tractament amb cèl·lules canceroses tractades amb el fàrmac. En el segon estudi es comprovarà quina resposta té el fàrmac a nivell poblacional, comparant el temps de supervivència d'individus tractats amb vemurafenib amb el d'individus tractats amb dacarbazine.

4.1. Estudi a nivell transcripcional de la inhibició de ^{V600E}BRAF en els melanomes

4.1.1. Presentació de l'estudi

Per dur a terme aquest anàlisi s'han utilitzat les dades pertanyents al treball realitzat per Parmenter et al^[31], les quals es poden descarregar lliurement del navegador web *Gene Expression Omnibus* (www.ncbi.nlm.nih.gov/geo)^[46]. En l'estudi de Parmenter et al, com en el que es presenta, es pretén avaluar l'efecte del vemurafenib a nivell transcripcional en cèl·lules de melanoma humà. Per fer-ho utilitzen la línia cel·lular A375, la qual presenta la mutació ^{V600E}BRAF, i incuben tres rèpliques tractades amb 10µM de vemurafenib i tres rèpliques control tractades amb 0.1% de DMSO (sis mostres en total) durant 24 hores. De cada mostra se n'obté el RNA resultant i se'n mesura l'expressió, en forma d'oligonucleòtids², mitjançant la hibridació amb sondes de microarrays³. Així, les dades obtingudes del servidor web GEO^[46] corresponen als valors d'expressió de cada sonda del microarray per a cada una de les sis mostres. Els

² Oligonucleòtid: seqüència de DNA o RNA de 50 parells de bases o inferior.

³ Microarray: superfície sòlida que conté una col·lecció d'*spots* de molècules de DNA, anomenades sondes, i que permet mesurar l'expressió gènica d'un determinat conjunt de gens a través de la hibridació amb les sondes.

microarrays utilitzats en l'estudi de Parmenter et al són del model Affymetrix Human Gene 1.0 ST Array, i els arxius d'expressió de les mostres venen en format CEL.

L'anàlisi es portarà a terme íntegrament amb RStudio, utilitzant principalment el paquet *oligo*^[40] de Bioconductor⁴ que permet l'anàlisi de dades de microarray. També s'han utilitzat altres paquets que ajuden a augmentar i complementar el procés, els quals s'aniran comentant durant l'explicació del procés seguit pel tractament de dades. Juntament amb la memòria s'adjunta l'script d'R utilitzat per dur a terme aquest anàlisi, el qual s'anomena **AnalisiMicroarray.R**.

Així, el que s'obtindrà al final d'aquest anàlisi serà una col·lecció de gens que es troben expressats de manera diferencial entre les dues condicions. És a dir, es valorarà aquells gens que es troben *up* i *down*-regulats (és a dir, sobreexpressats i infraexpressats/inhibits, respectivament) amb l'ús de vemurafenib respecte les cèl·lules control, relacionant-ho amb els processos biològics, funcions i vies metabòliques on estan implicats. Com l'objectiu és determinar si el vemurafenib frena la proliferació cel·lular en les cèl·lules canceroses, serà especialment important centrar l'atenció en aquells gens que es troben *down*-regulats en les cèl·lules tractades.

4.1.2. Desenvolupament i resultats de l'estudi

Tot i que existeixen diverses eines per a l'anàlisi de dades de microarray, en general totes elles segueixen un patró i tenen uns passos claus. En aquest apartat s'exposarà quin tractament han seguit les dades pas a pas, explicant per què i de quina manera s'ha dut a terme cada punt i quins són els resultats obtinguts. A més, s'inclouran les comandes que han permès seguir el procés i obtenir els resultats, però l'explicació de cada comanda s'inclou només a l'script adjunt per agilitzar la lectura de la memòria.

4.1.2.1. Lectura de les dades

El primer punt de l'estudi consisteix en la importació dels arxius CEL (descarregats prèviament del servidor web) a R i la creació de l'objecte de dades que contindrà els valors d'expressió de les sondes i amb el qual es treballarà. La importació dels arxius

⁴ Bioconductor: projecte per a l'anàlisi de dades genòmiques que conté paquets d'R per dur a terme a anàlisis de bioinformàtica i bioestadística^[47].

CEL (emmagatzemats en el directori *workingDir*) utilitzant el paquet *oligo* és un procediment senzill que consta dels dos següents passos:

```
> arxius_cel <- list.celfiles(workingDir, full.names=TRUE)
> dades <- read.celfiles(arxius_cel)
Platform design info loaded.
Reading in : C:/Users/Turias/Documents/Màster/Semestre 3/TFM/Dades/GSM1052615.CEL
Reading in : C:/Users/Turias/Documents/Màster/Semestre 3/TFM/Dades/GSM1052616.CEL
Reading in : C:/Users/Turias/Documents/Màster/Semestre 3/TFM/Dades/GSM1052617.CEL
Reading in : C:/Users/Turias/Documents/Màster/Semestre 3/TFM/Dades/GSM1052618.CEL
Reading in : C:/Users/Turias/Documents/Màster/Semestre 3/TFM/Dades/GSM1052619.CEL
Reading in : C:/Users/Turias/Documents/Màster/Semestre 3/TFM/Dades/GSM1052620.CEL
```

Les dades en cru (dades d'expressió de les sondes) s'han agrupat dins l'objecte *dades*, el qual conté els següents apartats:

```
> dades
GeneFeatureSet (storageMode: lockedEnvironment)
assayData: 1102500 features, 6 samples
  element names: exprs
protocolData
  rowNames: GSM1052615.CEL GSM1052616.CEL ... GSM1052620.CEL (6 total)
  varLabels: exprs dates
  varMetadata: labelDescription channel
phenoData
  rowNames: GSM1052615.CEL GSM1052616.CEL ... GSM1052620.CEL (6 total)
  varLabels: index
  varMetadata: labelDescription channel
featureData: none
experimentData: use 'experimentData(object)'
Annotation: pd.hugene.1.0.st.v1
```

Observant l'arxiu es pot comprovar que inclou les sis mostres (les tres rèpliques control i les tres rèpliques tractades) i 1102500 dades d'expressió de les sondes. S'ha de tenir en compte, però, que l'expressió d'una sonda no correspon a l'expressió d'un gen, ja que per cada gen hi ha diverses sondes que mesuren la seva intensitat. Així, tot aquest volum de dades es veurà reduït significativament en processos posteriors fins obtenir només aquells gens expressats diferencialment entre les dues condicions.

Tant els arxius CEL (mostres) com les dades d'expressió venen representats mitjançant un codi. En el primer cas, cada mostra té el codi *GSM10526XX.CEL*, on *XX* són números entre el 15 i el 20 depenent de la mostra. La informació per conèixer quins codis pertanyen a les mostres control i quines a les mostres tractades està inclosa en el mateix directori web d'on es poden descarregar^[45]. Així, els codis que acaben amb els números 15, 16 i 17 són les mostres control, mentre que les acabades amb 18, 19 i 20 són les tractades amb vemurafenib. En el segon cas, les sondes estan representades

per un codi numèric que més endavant s'haurà de traduir amb els noms dels gens corresponents, tot i que el procediment s'explicarà arribat el moment.

4.1.2.2. Exploració de les dades en cru

En aquest apartat s'han realitzat diversos gràfics que permeten observar la distribució de les expressions en les diferents mostres, com s'agrupen entre elles i, sobretot, la presència d'alguna mostra problemàtica. S'ha de tenir en compte que en aquest apartat es treballa amb totes les dades d'expressió de totes les sondes, així que no es poden treure més conclusions que comprovar que totes les mostres són de bona qualitat. Es realitzaran un total de set gràfics, que es presenten a continuació:

Histograma i boxplot

Aquests dos gràfics permeten observar com es distribueixen les expressions en les sis diferents mostres. L'histograma i el boxplot es mostren a la Figura 4.1 i Figura 4.2, respectivament, i la comanda utilitzada per crear-los es la següent:

```
> mostres <- c("Control_rep1","Control_rep2","Control_rep3",
+             "Treated_rep1","Treated_rep2","Treated_rep3")
> color <- c(rep("brown4",3),rep("cadetblue4",3))
> hist(dades,target="probeset",main="Histograma de les expressions",col=color)
> legend(x="topright",legend=mostres , col=color, cex=0.8, lty=1:6)
> boxplot(dades,target="probeset",col=color,las=1,
+         main="Boxplot de les expressions", cex=0.5, names=mostres)
.
```

Tant l'histograma com el boxplot mostren una certa uniformitat entre les rèpliques del mateix tractament, i no s'aprecia cap distribució anòmala en cap de les mostres.

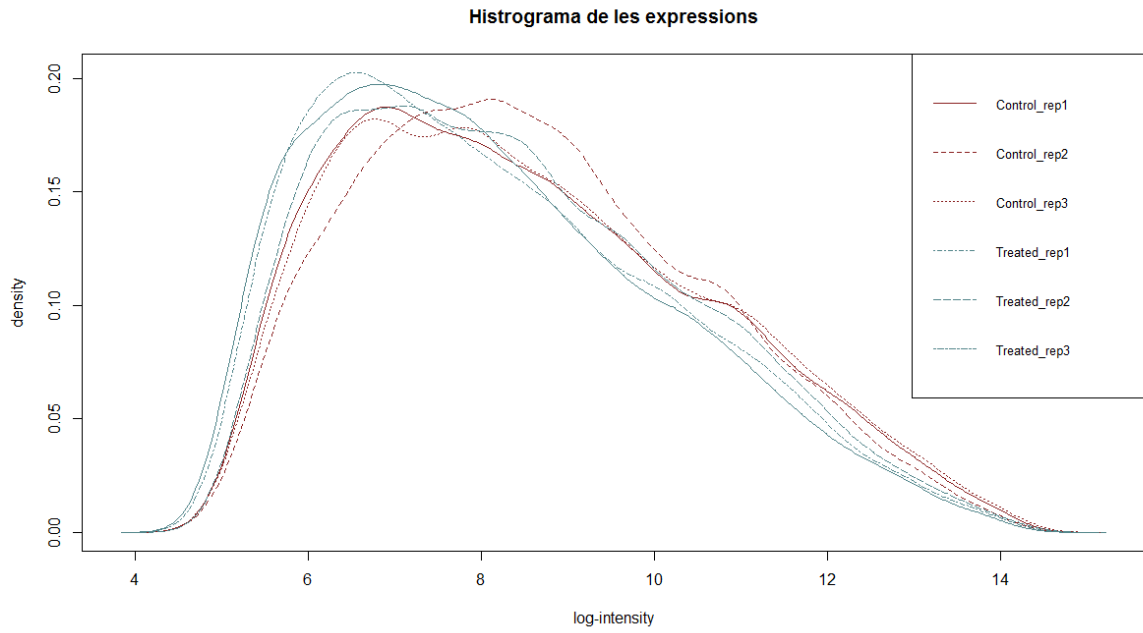


Figura 4.1. Histograma de la distribuci3 de les expressions de les sis mostres d'estudi.

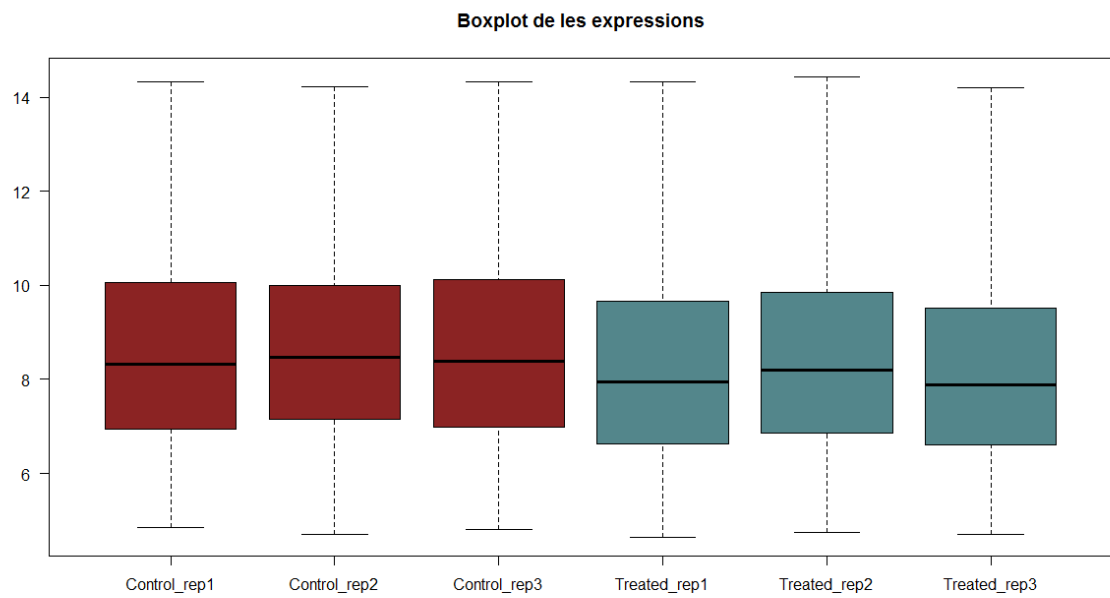


Figura 4.1. Boxplot de la distribuci3 de les expressions de les sis mostres d'estudi.

Clúster jeràrquic i anàlisi de components principals (PCA)

Aquests dos gràfics permeten observar com s'agrupen les mostres entre elles. En estar treballant amb mostres tractades i no tractades és d'esperar que les rèpliques s'agrupin seguint aquest criteri, tot i que si no és el cas no seria alarmant ja que, com s'ha explicat, s'esta utilitzant tot el conjunt de gens. El clúster jeràrquic es mostra a la Figura 4.3 i el PCA en la Figura 4.4, i la comanda utilitzada per crear-los és la següent:

```
> cluster <- hclust(dist(t(exprs(dades))),method="average")
> plot(cluster, labels=mostres, main="Agrupament jeràrquic de les mostres",
+       hang=-1)
>
> plotPCA <- function(X, labels=NULL, colors=NULL, dataDesc="", scale=FALSE)
+ {
+   pcX<-prcomp(t(X), scale=scale) # la funció 'prcomp()' permet dur a terme
+   # l'anàlisi de components principals
+   loads<- round(pcX$sdev^2/sum(pcX$sdev^2)*100,1)
+   xlab<-c(paste("PC1",loads[1],"%"))
+   ylab<-c(paste("PC2",loads[2],"%"))
+   if (is.null(colors)) colors=1
+   plot(pcX$x[,1:2],xlab=xlab,ylab=ylab, col=colors,
+        xlim=c(min(pcX$x[,1])-10, max(pcX$x[,1])+10))
+   text(pcX$x[,1],pcX$x[,2], labels, pos=3, cex=0.8)
+   title(paste("Gràfic dels dos primers PCs d'expressions de les mostres"),
+         cex=0.8)
+ }
> plotPCA(exprs(dades), labels=mostres)
```

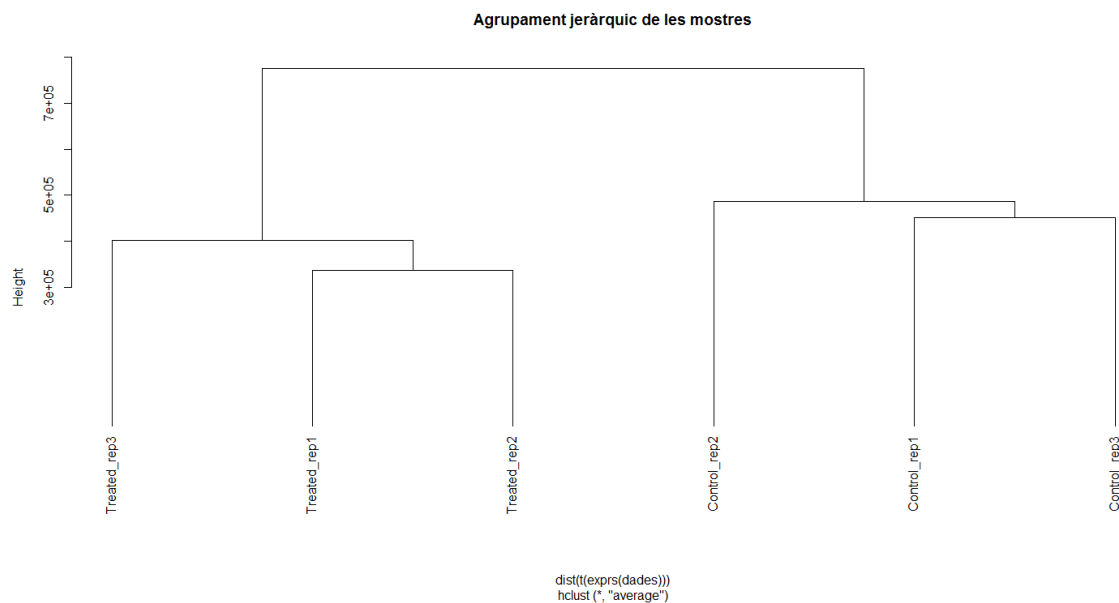


Figura 4.3. Agrupament jeràrquic de les sis mostres d'estudi.

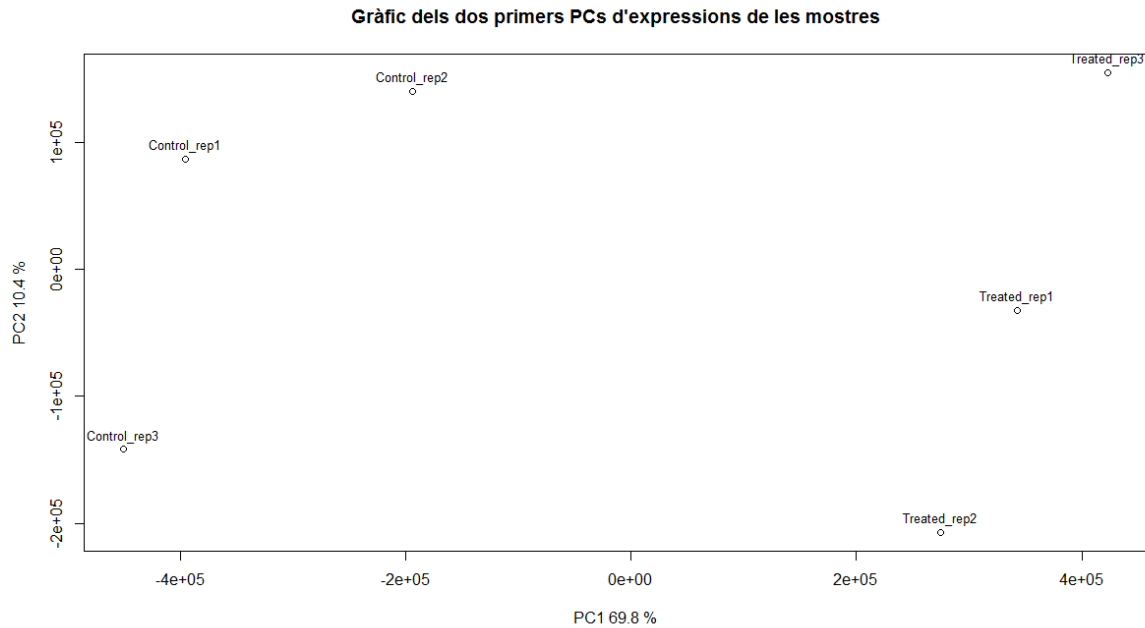


Figura 4.4. Gràfic d'anàlisi de components principals de les sis mostres d'estudi.

En ambdós casos les rèpliques tendeixen a agrupar-se segons si estan tractades amb vemurafenib o no.

MAplot

Un MAplot compara els valors d'expressió de cada mostra amb la mediana d'expressió de les sis mostres, permeten observar les desviacions presents en cada una de elles. La comanda per realitzar el MAplot, que es presenta a la Figura 4.5, és la següent:

```
> par(mfrow=c(2,3))
> MAplot(dades)
```

En els gràfics MAplot no s'aprecia una dispersió massa elevada en cap de les sis mostres.

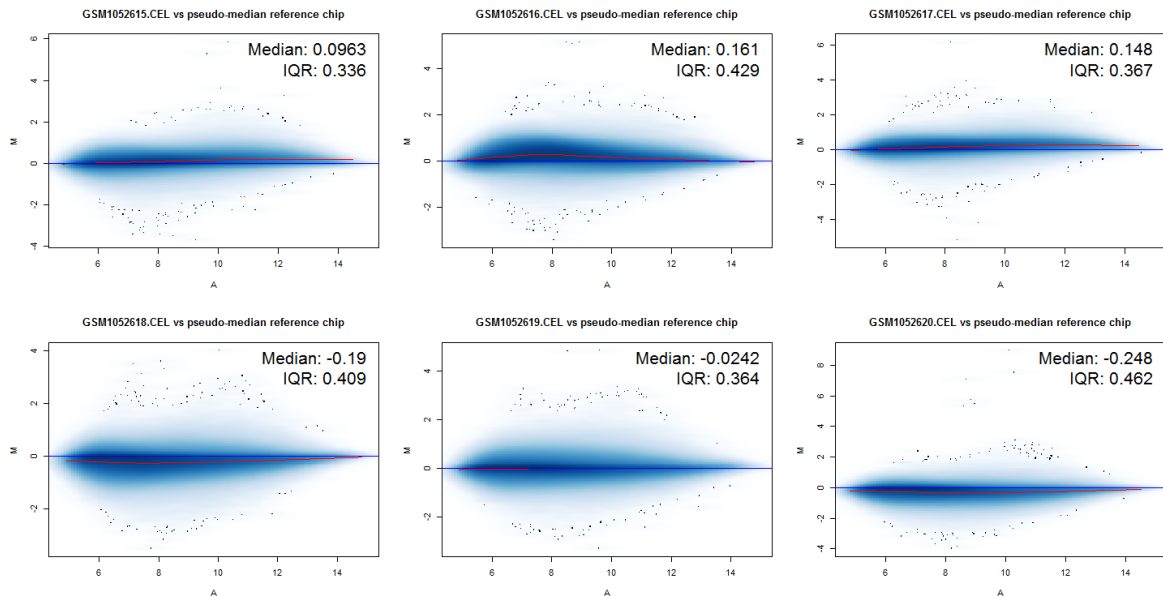


Figura 4.5. Gràfic MAplot de les sis mostres d'estudi.

RLE i NUSE

El gràfic d'expressió relativa en escala logarítmica (RLE, de l'anglès *Relative Log Expression*) i el gràfic d'error estandaritzat i normalitzat (NUSE, de l'anglès *Normalized Unscaled Standard Error*) permeten comprovar la presència d'errors en les mostres. Són gràfics tipus boxplot però per realitzar-los és necessari ajustar un model amb les dades d'expressió, essent els valors estimats del model comparats amb els valors reals^[47]. El més informatiu és el gràfic NUSE, ja que treballa amb els residus del model i permet detectar fàcilment problemes en els arrays de les mostres. Si alguna de les caixetes del boxplot no és simètrica i apareix desplaçada és indicatiu de problemes. Els gràfics RLE i NUSE es mostren a la Figura 4.6 i Figura 4.7, respectivament, i la comanda per realitzar-los és la següent:

```
> model0 <- fitProbeLevelModel(dades)
Background correcting... OK
Normalizing... OK
Summarizing... OK
Extracting...
  Estimates... OK
  StdErrors... OK
  Weights..... OK
  Residuals... OK
  Scale..... OK
> RLE(model0, main = "Expressió relativa (log): RLE", names=mostres,
+   las=1, col=color, ylim=c(-4,4))
> NUSE(model0, main = "Error no estandaritzat i normalitzat: NUSE",
+   las=1, names=mostres, las=2, col=color, ylim=c(0.5,3))
```

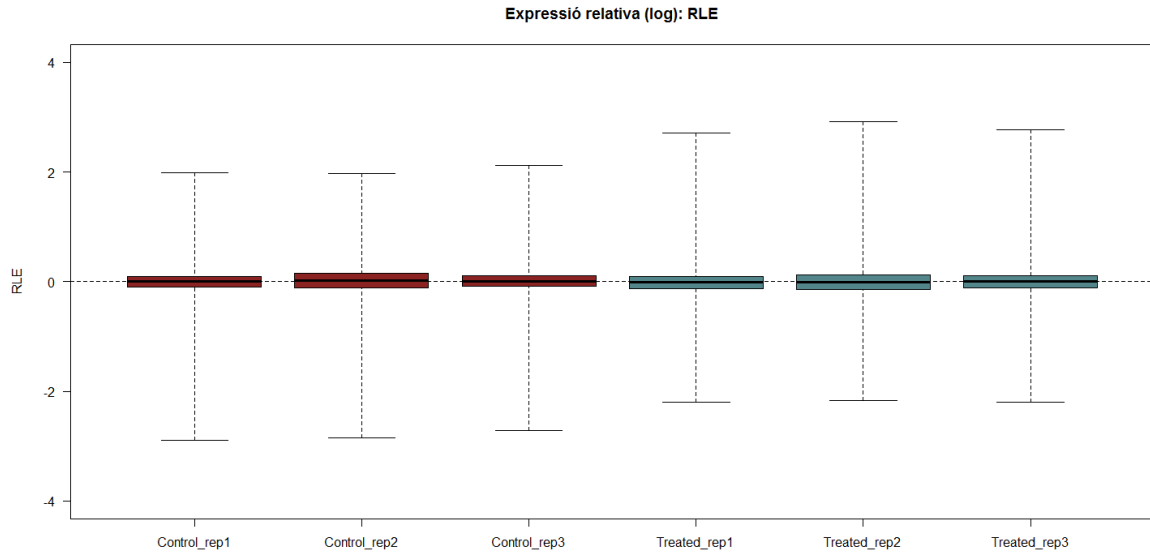


Figura 4.6. Gràfic d'expressió relativa (RLE) de les sis mostres d'estudi.

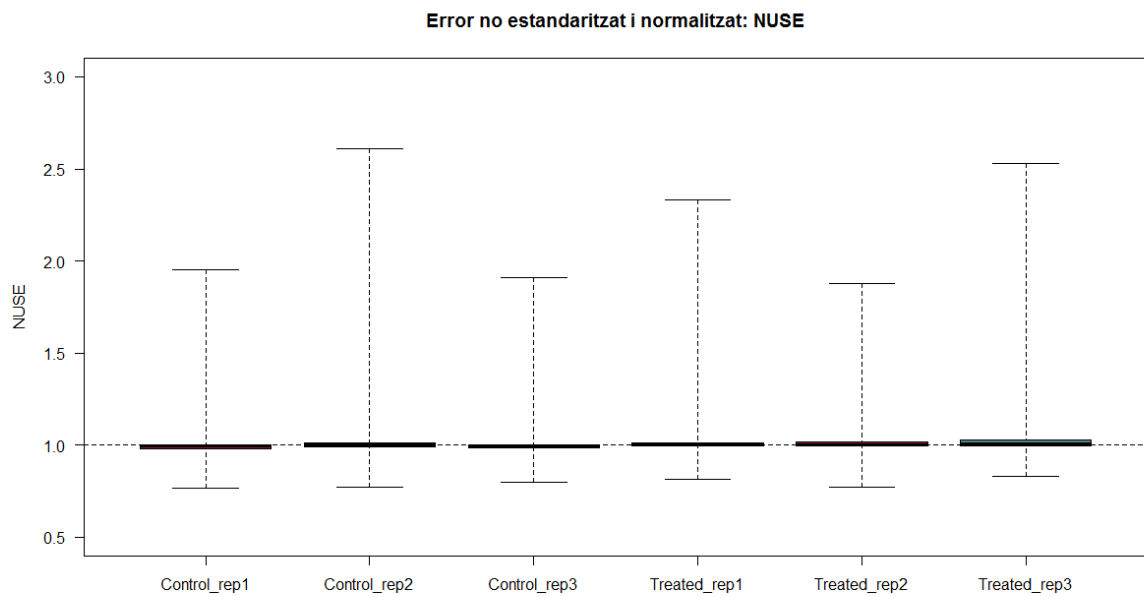


Figura 4.7. Gràfic d'error no estandaritzat i normalitzat (NUSE) de les sis mostres.

Tant el gràfic RLE com el NUSE són simètrics i centrats, així que no hi ha cap indicati que indiqui problemes en les mostres i, per tant, es pot procedir amb el següent pas de l'anàlisi.

4.1.2.3. Normalització de les dades

La normalització és una etapa important en molts anàlisis de dades de microarrays. Bàsicament, el que s'aconsegueix amb la normalització és ajustar tota aquella variància que hi pugui haver en les mostres, causada pels mètodes experimentals emprats, i fer les dades comparables entre sí^{[34][47]}.

Tot i que existeixen diversos mètodes de normalització, el més habitual entre arrays d'Affymetrix i el que s'ha utilitzat en aquest anàlisi és el mètode RMA. Aquest, a més de normalitzar les dades, prèviament realitza una correcció de fons i posteriorment a la normalització aplica una sumarització, és a dir, dóna un únic valor d'intensitat per a cada grup de sondes a partir de totes les sondes associades^[47]. En resum, un cop aplicat el mètode RMA s'aconsegueix que cada dada d'expressió correspongui a un gen i que siguin comparables entre elles.

Per dur a terme la normalització el mateix paquet *oligo* inclou l'opció *rma()*:

```
> dades_norm = oligo::rma(dades)
Background correcting
Normalizing
Calculating Expression
> matriu_dades = exprs(dades_norm)
> head(matriu_dades)
      GSM1052615.CEL GSM1052616.CEL GSM1052617.CEL GSM1052618.CEL GSM1052619.CEL GSM1052620.CEL
7892501      7.245602      6.806879      7.733020      6.189622      7.053367      7.203720
7892502      6.827139      6.701600      7.024739      6.204961      6.765573      6.242556
7892503      4.399835      4.507868      4.882556      4.363014      4.181428      4.734985
7892504      9.480247      9.679510      9.630740      9.691996      9.913263      9.658967
7892505      4.547398      4.452535      5.117582      4.873131      5.155099      3.993458
7892506      6.807030      6.905987      6.724742      6.770616      6.770599      6.776866
> dim(matriu_dades)
[1] 33297      6
> dim(exprs(dades))
[1] 1102500    6
```

La matriu *matriu_dades* conté tots els valors d'expressió de les sis mostres normalitzades i, com es pot observar, el número d'aquests s'ha reduït dràsticament, passant de 1102500 a 33297. A més, com s'ha explicat anteriorment i es pot observar, cada gen ve anotat mitjançant un codi numèric que, posteriorment, s'haurà d'identificar.

Exploració gràfica de les dades normalitzades

Per acabar aquest apartat, s'ha dut a terme una altra exploració gràfica per tal de visualitzar el resultat de la normalització. No obstant això, en aquets cas només s'ha realitzat un boxplot (Figura 4.8), un MAplot (Figura 4.9) i un gràfic PCA (Figura 4.10).

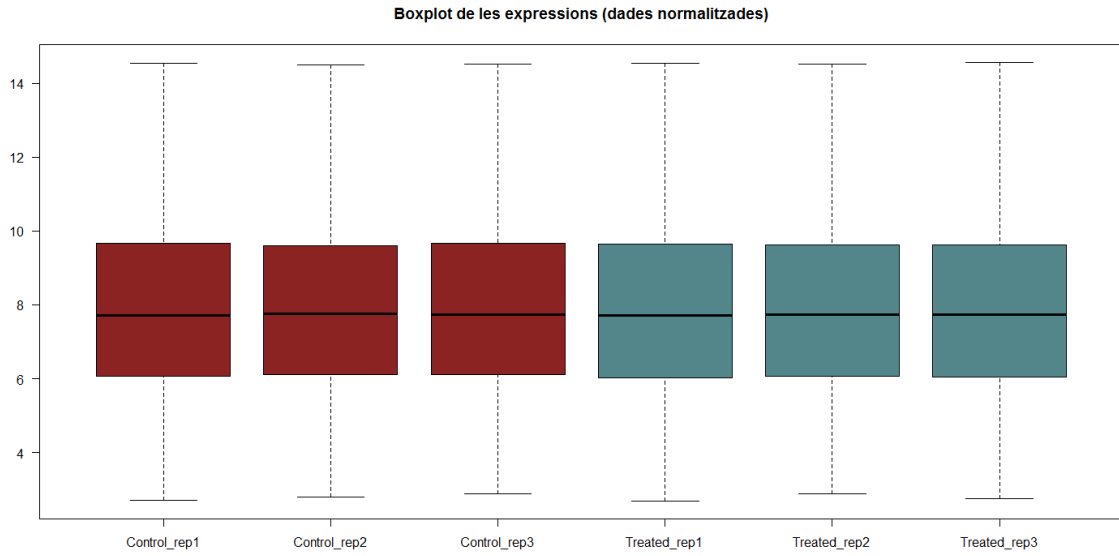


Figura 4.8. Boxplot de la distribució de les expressions de les sis mostres normalitzades.

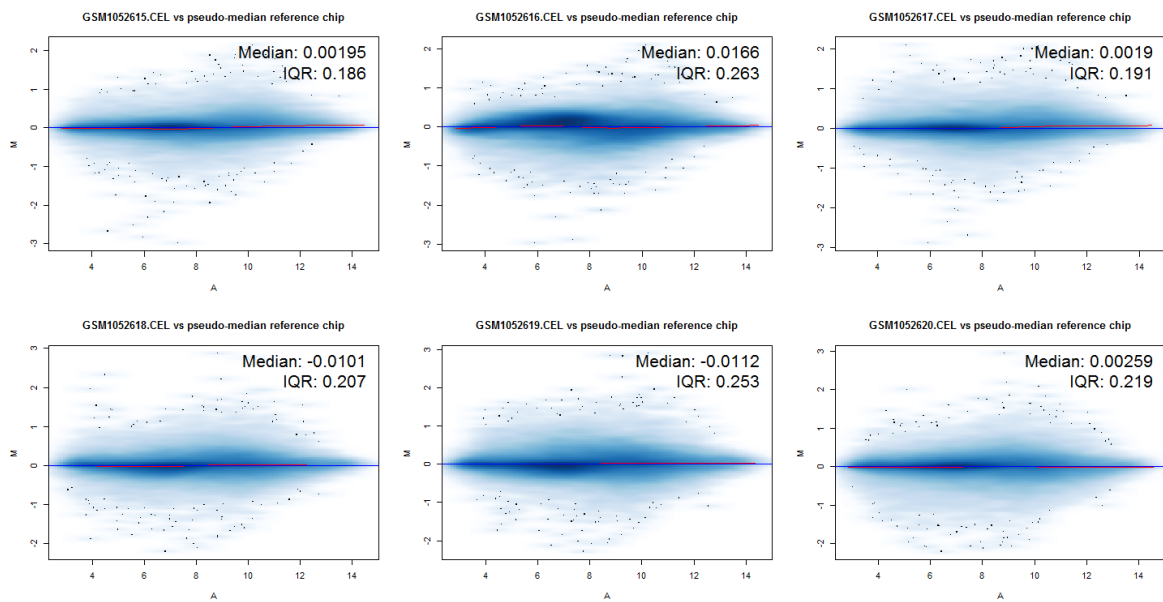


Figura 4.9. Gràfic MAplot de les sis mostres normalitzades.

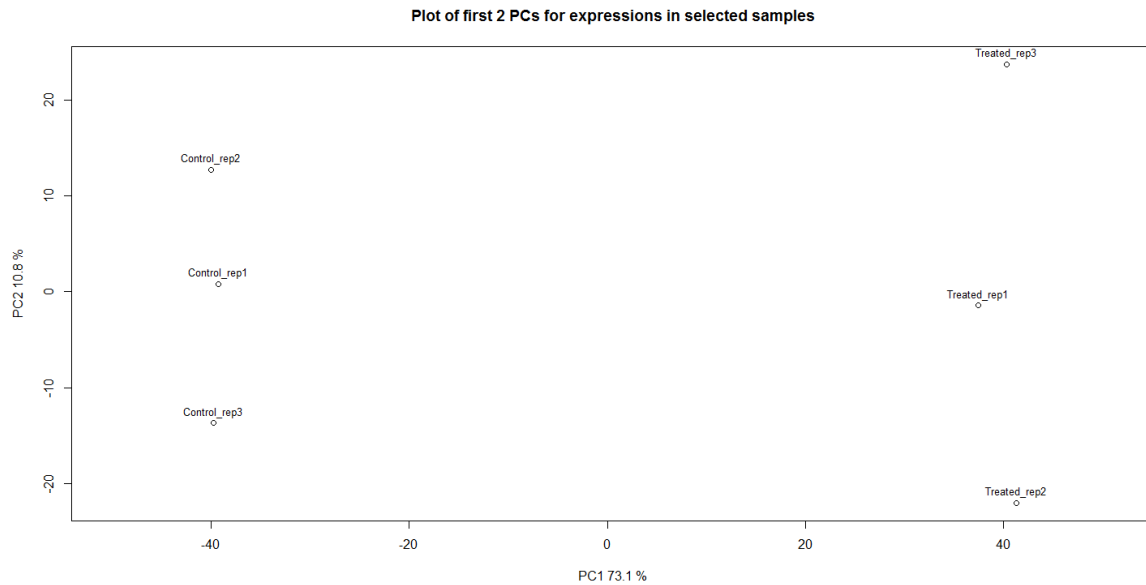


Figura 4.10. Gràfic d'anàlisi de components principals de les sis mostres normalitzades.

Com s'aprecia en el gràfics, l'efecte de la normalització de les dades és clara. En el boxplot s'observa com la distribució de les expressions en les sis mostres s'ha pràcticament igualat, mentre que en el MAplot la dispersió és molt més homogènia entre les mostres. Finalment, en el gràfic de components principals les rèpliques presenten una major agrupació segons si han estat tractades o no.

4.1.2.4. Selecció de gens expressats diferencialment

Un cop les dades estan normalitzades es pot procedir amb la selecció de gens. Per a la selecció dels gens diferencialment expressats existeixen múltiples aproximacions que permeten dur-ho a terme. En aquest cas s'ha usat el mètode basat en la creació d'un model lineal, a partir del qual es podrà determinar aquells gens que difereixen significativament entre les dues condicions (control i tractament). El paquet *limma*^[35] proporciona les eines adequades per realitzar el model lineal.

Creació i estimació del model lineal

Per a l'estimació del model lineal és necessari, en primer lloc, la creació d'una matriu de disseny i una matriu de contrast. Per una banda, una matriu de disseny consisteix en una matriu d'uns i zeros que determina quina mostra pertany a cada tractament. Les mostres s'assignen a les files i els tractaments a les columnes, i la matriu s'omple d'uns i zeros segons si la mostra pertany o no al tractament. Per altra banda, la matriu

de contrast és la que permet realitzar les comparacions, que en aquest cas serà la comparació entre les mostres amb vemurafenib i les mostres control.

La matriu de disseny es pot crear de manera manual o utilitzant la funció `model.matrix()` d'R:

```
> f = factor(rep(c("control","vemurafenib"),each=3),
+           levels=c("control","vemurafenib"))
> disseny = model.matrix(~ 0 + f)
> colnames(disseny) = c("control","vemurafenib")
> disseny
  control vemurafenib
1        1           0
2        1           0
3        1           0
4        0           1
5        0           1
6        0           1
attr(,"assign")
[1] 1 1
attr(,"contrasts")
attr(,"contrasts")$f
[1] "contr.treatment"
```

La matriu de contrast es crea a partir de la comanda `makeContrast()`, incloent-hi la comparació que es desitja realitzar (`vemurafenib – control`):

```
> matriu_contrast = makeContrasts(efecte=(vemurafenib-control),levels=disseny)
> matriu_contrast
      Contrasts
Levels   efecte
control   -1
vemurafenib  1
```

Com s'observa, el resultat és l'assignació del valor "-1" al control i "1" al vemurafenib. Un cop creades la matriu de disseny i la matriu de contrast ja és possible estimar el model lineal, la creació del qual és un procés de tres etapes.

La primera etapa és l'estimació d'un model utilitzant només la matriu de dades (`matriu_dades`, obtinguda a l'apartat anterior) i la matriu de disseny, ja que d'aquesta manera es calcula la mitjana d'expressió de les tres rèpliques per a cada gen i tractament. El segon pas és aplicar la matriu de contrast al model creat, comparant així els nivells d'expressió entre mostres control i les mostres amb vemurafenib per a cada gen. Finalment, s'han d'utilitzar mètodes empírics Bayesianos per tal de reduir la variància del model, és a dir, per reduir la influència de desviacions atípiques molt altes o molt baixes. Los comandes són les següents:

```

> model = lmFit(matriu_dades,disseny)
> model_contrast = contrasts.fit(model,matriu_contrast)
> model_eb = eBayes(model_contrast)

```

Així, el model *model_eb* final conté, entre altres, els coeficients d'expressió dels gens, el contrast aplicat i valors t i p-valors que permetran seleccionar els gens expressats de manera diferencial. Una manera de visualitzar els resultats del model és mitjançant un volcano plot (Figura 4.11), el qual representa els canvis d'expressió dels gens entre les dues condicions. La comanda és la següent:

```

> volcanoplot(model_eb,coef=1, highlight=10, names=model_eb$ID,
+             main=paste("Expressió de gens diferencials",
+             colnames(matriu_contrast)[1], sep="\n"))
> abline(v=c(-1,1))

```

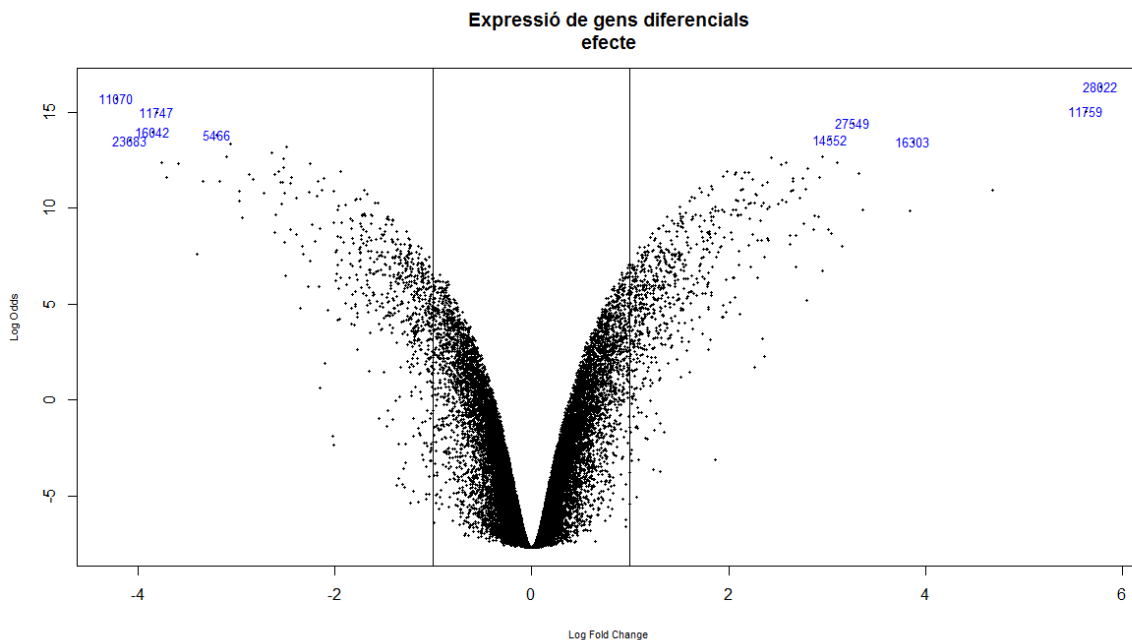


Figura 4.11. Volcano plot dels gens expressats. Els valors superior a (-1,1) a l'eix de les X (*Log Fold Change*) i superiors a 0 a l'eix de les Y (*Log Odds*) són els que es troben diferencialment expressats.

Aquells gens que superen el valor de 1, en valor absolut, a l'eix de les X i el valor de 0 a l'eix de les Y són els que es troben diferencialment expressats.

Selecció de gens

Un cop estimat el model es pot procedir a la selecció dels gens que es troben expressats de manera diferencial entre les dues condicions. Treballant amb una significació del 5%, la funció *decideTests()* del paquet *limma* evalua cada gen del model

creat i identifica aquells que són significativament diferents sota la hipòsis nula següent:

H_0 : no hi ha diferències significatives en l'expressió genètica.

El resultat d'aplicar aquesta funció és tota la llista de gens els quals tindran assignats un valor de tres possibles: -1 els *down*-regulats, 1 els *up*-regulats i 0 si no s'aprecien diferències significatives. És a dir, aquells gens que augmenten la seva expressió significativament per l'efecte de vemurafenib se'ls assignarà el número 1, mentre que aquells que disminueixen l'expressió (o són inhibits) pel vemurafenib se'ls assignarà el número -1. La comanda és la següent:

```
> gens <- decideTests(model_eb,method='global',adjust.method="BH", p.value=0.05,
+                       lfc=1)
> gens[1:50,]
7892501 7892502 7892503 7892504 7892505 7892506 7892507 7892508 7892509 7892510
      0      0      0      0      0      0      0      0      0      0
7892511 7892512 7892513 7892514 7892515 7892516 7892517 7892518 7892519 7892520
      0      0      0      0      0      0      0      0      0      0
7892521 7892522 7892524 7892525 7892526 7892527 7892528 7892529 7892530 7892531
      0      0      0      0      -1      0      0      0      0      0
7892532 7892533 7892534 7892535 7892536 7892537 7892538 7892539 7892540 7892541
      0      0      0      0      0      0      0      0      0      0
7892542 7892543 7892544 7892545 7892547 7892548 7892549 7892550 7892551 7892552
      0      0      0      0      0      0      0      0      0      0
```

La quantitat de gens *up* i *down*-regulats per l'efecte del vemurafenib és la següent:

```
> print(summary(gens))
efecte
-1     604
 0    3192
 1     701
```

Com s'observa, la quantitat de gens que varien la seva expressió és mínima respecte els que no la varien, però són aquests els gens d'interès i els que donaran resposta a quin efecte produeix el vemurafenib en les cèl·lules de melanoma humà. Així, el següent pas és seleccionar només els gens que es troben diferencialment expressats de tota la llista de gens:

```
> gens_DE <- gens[gens!=0,]
> df_gens_DE <- data.frame(gens_DE)
> head(df_gens_DE)
      gens_DE
7892526    -1
7892570    -1
7892609    -1
7892674    -1
7892676    -1
7892702    -1
```


Amb això ja s'ha aconseguit un *data frame* només d'aquells gens diferencialment expressats. Un heatmap (Figura 4.12) permet visualitzar els perfils d'expressió dels gens seleccionats, aplicant diferents colors segons si els gens es troben *up* o *down*-regulats. La comanda per la creació del heatmap és la següent:

```
> hm<-exprs(dades_norm)[ID_gens_DE,]  
> heatmap(hm,cexCol=0.9)
```

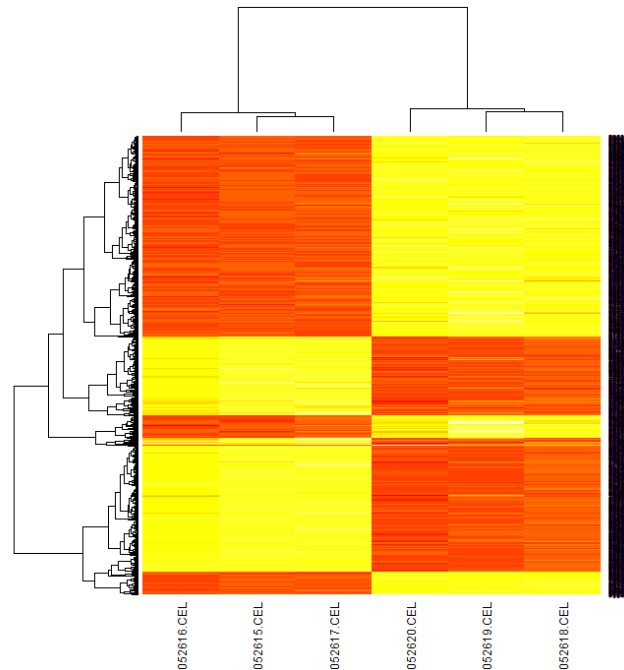


Figura 4.12. Heatmap dels gens diferencialment expressats entre les mostres tractades amb vemurafenib i les mostres control.

Com s'observa a la Figura 4.12, la diferència d'intensitat en l'expressió és clara entre les dues condicions.

4.1.2.5. Anotació de resultats

Tot i haver seleccionat els gens que es troben diferencialment expressats, realment serveix de poc si no es coneix quins són aquests gens. Com s'ha comentat anteriorment, tots els gens provinents dels microarrays venen anotats mitjançant un codi numèric. Per traduir aquest codi i identificar el gen al qual estan associats existeixen una sèrie de paquets d'anotació que depenen del microarray utilitzat. El paquet que s'ha utilitzat en aquest cas per anotar els gens és el *hugene10sttranscriptcluster.db*^[54]. A més, també s'ha utilitzat el paquet *annaffy*^[55], el qual permet la creació de taules d'anotació on apareix informació de cada gen anotat com el símbol, el nom i el cromosoma en el qual està localitzat, així com hipervincles a

diverses bases de dades amb informació relativa a cada gen. Així, s'han creat dues taules d'anotació, una pels gens *up*-regulats i una altra pels *down*-regulats, anomenades **AnotacioGensUp.html** i **AnotacioGensDown.html**, respectivament, i que s'entreguen juntament amb la memòria.

Per crear aquestes taules és necessari, primer, emmagatzemar els codis dels gens en dos objectes que s'han anomenat *ID_up* i *ID_down*:

```
> ID_up <- rownames(subset(df_gens_DE, df_gens_DE[,1]==1))
> ID_down <- rownames(subset(df_gens_DE, df_gens_DE[,1]==-1))
> ID_gens_DE<-rownames(df_gens_DE)
> head(ID_up)
[1] "7892850" "7893862" "7895856" "7896727" "7897441" "7897675"
```

L'objecte *ID_gens_DE* conté la ID de tots els gens diferencialment expressats, però aquest serà necessari en el següent apartat. Un cop creats els nous objectes s'ha procedit a la creació de les dues taules d'anotació dels gens, essent un exemple del procés el següent:

```
> taula_down<-aafTableAnn(as.character(ID_down), "hugene10sttranscriptcluster.db")
> saveHTML(taula_down, file.path(workingDir, "AnotacioGensDown.html"),
+         "Anotació dels gens down regulats")
```

En els arxius obtinguts apareixen gens que no estan anotats així com gens en els quals els falta alguna informació, però en general les taules són bastant completes. No obstant això, treure conclusions sobre l'efecte del fàrmac vemurafenib d'aquest arxiu és una tasca tediosa i ineficient, ja que s'hauria de revisar gen a gen les funcions que desenvolupen i relacionar-ho manualment. És per aquest motiu que un anàlisi de significació biològica és vital en aquests tipus d'estudis.

4.1.2.6. Anàlisi de significació biològica

Fins aquest punt s'ha obtingut tota una col·lecció de gens que es troben expressats de manera diferencial per l'efecte del vemurafenib, ja sigui per sobreexpressió o per inhibició, però són poc informatiu per si sols, i per aquest motiu és clau un anàlisi de significació biològica. De forma resumida, l'anàlisi de significació biològica elabora una llista dels termes GO i termes KEGG més comuns entre tota aquesta col·lecció de gens seleccionats.

Gene Ontology (GO)^{[48][49]} i *Kyoto Encyclopedia of Genes and Genomes* (KEGG)^{[50][51][52]} són bases de dades que donen informació sobre les funcions i processos biològics,

entre altres, dels gens anotats. Els termes GO informen de les funcions que desenvolupen els gens, els processos biològics on estan implicats i el component cel·lular on actuen, mentre que els termes KEGG donen una informació més global de la funció del gen dins la cèl·lula. Així, l'anàlisi de significació biològica consisteix en relacionar cada gen seleccionat amb els termes GO i KEGG que li corresponguin, per després elaborar una llista d'aquells termes més repetits. D'aquesta manera es podrà determinar com afecta a nivell funcional i de processos la presència del vemurafenib en les cèl·lules canceroses i comprovar si inhibeix o no la proliferació cel·lular.

Mitjançant els paquets *GOstats*^[56], *hugene10sttranscriptcluster.db* i *org.Hs.eg.db*^[57] és possible vincular els gens amb els servidors web *Gene Ontology* i KEGG. Els gens s'han separats segons si són *up* o *down*-regulats, fet que permetrà comprovar quins processos estan sobreexpressats i quins inhibits per l'efecte del vemurafenib. Així, d'aquest anàlisi s'obtindran quatre arxius html que s'adjunten amb la memòria:

- **GOup**: termes GO dels gens *up*-regulats.
- **GOdown**: termes GO dels gens *down*-regulats.
- **KEGGup**: termes KEGG dels gens *up*-regulats.
- **KEGGdown**: termes KEGG dels gens *down*-regulats.

No obstant això, com que es pretén avaluar si el vemurafenib és un fàrmac eficaç per aturar la proliferació cel·lular en cèl·lules canceroses, és lògic esperar que on es trobarà més informació d'interès és en els gens *down*-regulats. A més, com s'ha explicat, dels termes GO se'n pot obtenir la funció molecular, el procés biològic i el component cel·lular, però en aquest projecte només s'avaluaran els termes GO corresponents al procés biològic. Tenint en compte que es vol comprovar l'efecte del vemurafenib sobre el procés de proliferació cel·lular, conèixer les funcions moleculars i els components cel·lulars més comuns no aportaria informació rellevant ni necessària. Un exemple del procés per obtenir aquests quatre arxius és el següent:

```

> for (i in 1:length(l1listat)){
+   geneIds_up<-unique(getEG(as.character(ID_up),"hugene10sttranscriptcluster.db"))
+   GOparams_up = new("GOHyperGParams",
+                     geneIds=geneIds_up,
+                     universeGeneIds=entrez,
+                     annotation="org.Hs.eg.db",
+                     ontology="BP",
+                     pvalueCutoff=0.05,
+                     conditional=FALSE,
+                     testDirection="over")
+   KEGGparams_up = new("KEGGHyperGParams",
+                       geneIds=geneIds_up,
+                       universeGeneIds=entrez,
+                       annotation="org.Hs.eg.db",
+                       pvalueCutoff=0.05,
+                       testDirection="over")
+   GOhyper_up = hyperGTest(GOparams_up)
+   KEGGhyper_up = hyperGTest(KEGGparams_up)
+   GOfilename_up=file.path(workingDir,
+                           paste("GOup.html"))
+   KEGGfilename_up=file.path(workingDir,
+                              paste("KEGGup.html"))
+   htmlReport(GOhyper_up,file=GOfilename_up,
+              summary.args=list("htmlLinks"=TRUE))
+   htmlReport(KEGGhyper_up,file=KEGGfilename_up,
+              summary.args = list("htmlLinks"=TRUE))
+ }

```

A la taula obtinguda dels termes GO *down*-regulats (**GOdown**) hi destaquen els processos que tenen relació amb el desenvolupament i la divisió cel·lular (cicle cel·lular, mitosi, replicació del DNA, etc.) així com processos que involucren el DNA (reparació, organització dels cromosomes, etc.). La llista de termes KEGG *down*-regulats (**KEGGdown**) apunta en la mateixa direcció, essent les funcions i vies relacionades amb el cicle cel·lular les que presenten més diferències d'expressió respecte les cèl·lules control.

Pel que fa els termes GO i KEGG *up*-regulats (**GOup** i **KEGGup**), en el primer cas apareixen processos diversos relacionats amb el transport, locomoció, processos metabòlics de molècules petites i migració cel·lular. En el cas dels termes KEGG també apareixen funcions i vies diverses, tot i que s'ha de destacar la presència d'un augment en l'expressió de gens relacionats amb les vies MAPK.

4.2. Estudi a nivell poblacional de la inhibició de ^{V600E}BRAF en els melanomes

4.2.1. Presentació de l'estudi

Per dur a terme aquest anàlisi s'han utilitzat les dades de l'estudi realitzat per Chapman et al^[41], en el qual es pretén avaluar i comparar, entre altres, la supervivència d'individus tractats amb vemurafenib amb la d'individus tractats amb dacarbazine. Com s'ha explicat anteriorment, el dacarbazine és un medicament usat també per combatre els melanomes però que presenta una baixa eficàcia^[30]. En l'estudi es treballa amb 672 individus que presenten la mutació ^{V600E}BRAF en melanomes, els quals se separen en dos grups d'igual número (336). Un grup és tractat amb vemurafenib i un altre grup amb dacarbazine, i es realitza un seguiment durant un any de la supervivència i mort d'aquests individus.

En aquest apartat, doncs, es durà a terme un anàlisi de supervivència⁵ amb l'objectiu d'avaluar si els individus tractats amb vemurafenib presenten una major supervivència respecte els individus tractats amb dacarbazine, comprovant si la diferència observada (en cas d'haver-n'hi) és significativa o no. A l'hora de realitzar un anàlisi de supervivència s'han de tenir en compte dos factors: el temps de falla o temps fins l'esdeveniment i el temps de censura. En aquest cas, l'esdeveniment o falla és la mort de l'individu, mentre que la censura és quan s'interromp el seguiment d'un individu per una causa diferent a la falla (per exemple, que abandona l'estudi). Per tant, de cada individu se n'ha de mesurar el temps des de l'inici de l'estudi fins que falla, és censurat o fins la finalització del projecte.

Aquest anàlisi s'inicia, doncs, amb els temps de falla i censura dels individus, i a partir d'aquests es durà a terme l'anàlisi de supervivència. No obstant això, les dades exactes dels temps de falla i censura no s'han pogut obtenir ja que no estan disponibles de forma lliure. Així, s'ha realitzat una aproximació dels temps de falla i censura a partir del gràfic de les corbes de supervivència de l'estudi, el qual es mostra a la Figura 4.13. A l'apartat 4.2.2.1 de creació de les dades es donaran detalls de com s'han obtingut.

⁵ Anàlisi de supervivència: col·lecció de procediments estadístics on la variable d'interès és el temps fins que es produeix un esdeveniment (mort, recuperació del pacient, etc.)^[42].

A Overall Survival

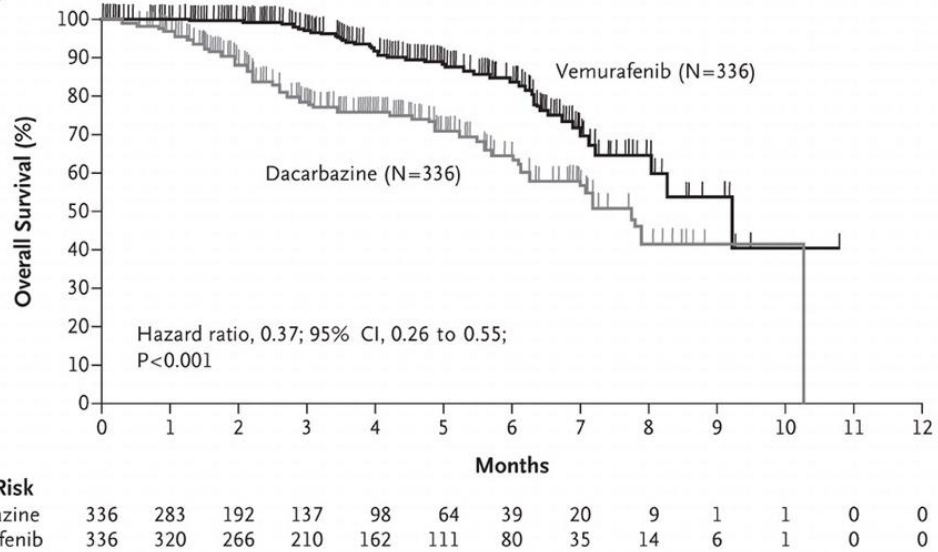


Figura 4.13. Gràfic de les corbes de supervivència pels individus tractats amb vemurafenib (línia negra) i dacarbazine (línia grisa) obtingut de l'estudi dut a terme per Chapman et al^[41]. Les línies verticals de les corbes representen els individus censurats durant el temps d'estudi. A l'eix de les X es mostra el temps de supervivència (mesos) i a l'eix de les Y el percentatge de supervivència.

El que s'espera al final d'aquest estudi és obtenir una corba de supervivència per a cada grup de tractament, semblants a les corbes de la Figura 4.13, i contrastar la següent hipòtesis nul·la:

H_0 : no hi ha diferències significatives entre les corbes de supervivència (entre tractaments).

L'anàlisi es portarà a terme íntegrament amb RStudio, utilitzant només el paquet *survival*^[53] inclòs a R i que presenta totes les eines i funcions necessàries per realitzar l'anàlisi de supervivència. Amb la memòria s'adjunta un script d'R amb el procediment seguit en l'anàlisi de supervivència i anomenat **AnalisiSupervivencia.R**.

4.2.2. Desenvolupament i resultats de l'estudi

Un anàlisi de supervivència no té la complexitat d'un anàlisi de dades de microarray però sí que ha de seguir una sèrie de passos. En aquest apartat s'exposarà quin tractament han seguit les dades pas a pas, explicant per què i de quina manera s'ha dut a terme cada punt i quins són els resultats obtinguts. A més, s'inclouran les comandes que han permès seguir el procés i obtenir els resultats, però l'explicació de cada comanda s'inclou només a l'script adjunt per agilitzar la lectura de la memòria.

4.2.2.1. Creació de les dades per l'estudi

El primer pas és crear els objectes de dades per a realitzar l'anàlisi de supervivència, essent només necessaris els temps de falla i els temps de censura dels individus. Com s'ha explicat, aquests temps no es coneixen de manera exacte i s'han aproximat a partir dels individus que han fallat i han estat censurats cada mes. Per dur-ho a terme primer s'han comptabilitzat els individus censurats cada mes mitjançant el gràfic de la Figura 4.13, i a partir d'aquests i el número de persones en risc (*No. at risk*) s'han calculat els individus que han fallat de la següent manera:

$$N^{\circ} \text{ d'esdeveniments} = (N^{\circ} \text{ en risc}_{\text{mes } i} - N^{\circ} \text{ en risc}_{\text{mes } i+1}) - N^{\circ} \text{ de censurats}$$

A la Taula 4.1 i Taula 4.2 es mostren el número d'individus en risc, els individus que han fallat i els censurats cada mes per a cada tractament.

Taula 4.1. Recompte d'individus que han seguit el tractament amb vemurafenib mes a mes durant un any.

Vemurafenib			
	Número individus en risc (a l'inici del mes)	Número esdeveniments (durant tot el mes)	Número censurats (durant tot el mes)
Gener	336	0	16
Febrer	320	1	53
Març	266	21	35
Abril	210	31	17
Maig	162	26	25
Juny	111	15	16
Juliol	80	21	24
Agost	35	8	13
Setembre	14	2	6
Octubre	6	1	4
Novembre	1	0	1
Desembre	0	0	0
Final	0	0	0

Taula 4.2. Recompte d'individus que han seguit el tractament amb dacarbazine mes a mes durant un any.

Dacarbazine			
	Número individus en risc (a l'inici del mes)	Número esdeveniments (durant tot el mes)	Número censurats (durant tot el mes)
Gener	336	21	32
Febrer	283	51	40
Març	192	39	16
Abril	137	13	26
Maig	98	14	20
Juny	64	11	14
Juliol	39	9	10
Agost	20	4	7
Setembre	9	0	8
Octubre	1	0	0
Novembre	1	1	0
Desembre	0	0	0
Final	0	0	0

Un cop calculats els individus fallats i censurats mes a mes, els temps de falla i censura s'han obtingut repartint el número d'individus equitativament al llarg del mes. Per exemple, si durant el mes de gener s'han censurat 16 individus que seguien el tractament amb vemurafenib, els temps de censura s'han obtingut de la següent manera:

$$Censurats\ gener = \frac{1\ mes}{16\ censurats} \cdot \frac{31\ dies}{1\ mes} = 1.94\ dia/censurat$$

Així, durant el mes de gener s'ha considerat que hi ha un individu censurat cada 1.94 dies. Totes les dades s'han calculat a mà, i òbviament en els mesos on només hi havia un o dos individus no ha fet falta seguir aquest procediment. Un cop calculats els temps de falla i censura s'han introduït dins d'R, un exemple del qual és el següent:

```
> esd.v <- round(c(1.5,seq(2,2.96,by=0.0476),seq(3,3.97,by=0.0322),
+               seq(4,4.97,by=0.038),seq(5,5.94,by=0.066),
+               seq(6,6.96,by=0.0476),seq(7,7.875,by=0.125),9.25,9.25,10.25),2)
> cens.v <- round(c(seq(0,0.94,by=0.0625),seq(1,1.99,by=0.0189),
+                 seq(2,2.98,by=0.0286),seq(3,3.942,by=0.0588),
+                 seq(4,4.96,by=0.04),seq(5,5.94,by=0.0625),
+                 seq(6,6.97,by=0.042),seq(7,7.93,by=0.0769),
+                 seq(8,8.84,by=0.167),seq(9,9.75,by=0.25),10.75),2)
> temps.v <- c(esd.v,cens.v)
```

Els objectes *esd.v* i *cens.v* contenen els temps de falla i censura, respectivament, pels individus tractats amb vemurafenib, mentre que *temps.v* conté el temps de falla i

censura junts. S'ha procedit d'igual manera amb els individus tractats amb dacarbazine, obtenint els objectes *esd.d*, *cens.d* i *temps.d*. Òbviament només amb aquest conjunt de temps no és possible dur a terme l'anàlisi, ja que R no pot identificar quins temps corresponen a falles i quins a censures ni quins són els individus tractats amb vemurafenib i dacarbazine. Per tant, és necessària la creació de noves variables que relacionin aquests conceptes. El procediment seguit ha estat el següent:

```
> length(esd.v)
[1] 126
> length(cens.v)
[1] 210
> length(esd.d)
[1] 163
> length(cens.d)
[1] 173
> tractament <- rep(c("V","V","D","D"),times=c(126,210,163,173))
> status <- c(rep(1,times=126),rep(0,times=210),rep(1,times=163),rep(0,times=173))
> temps <- c(temps.v,temps.d)
```

La variable *tractament* indica quins temps corresponen a cada tractament ("V" per al tractament amb vemurafenib i "D" per al tractament amb dacarbazine), i la variable *status* diferencia entre els temps de falla, que corresponen al valor "1", i els temps de censura, que corresponen al valor "0". Finalment, la variable *temps* engloba tots els temps junts, i amb aquests objectes s'ha creat el *data frame* incloent totes les dades i amb el qual es pot dur a terme l'anàlisi de supervivència:

```
> dades2 <- data.frame(temps,tractament,status)
> head(dades2)
  temps tractament status
1  1.50          V      1
2  2.00          V      1
3  2.05          V      1
4  2.10          V      1
5  2.14          V      1
6  2.19          V      1
> tail(dades2)
  temps tractament status
667  8.25          D      0
668  8.38          D      0
669  8.50          D      0
670  8.62          D      0
671  8.75          D      0
672  8.88          D      0
```

4.2.2.2. Càlcul de l'estimador de Kaplan-Meier

L'estimador de Kaplan-Meier o estimador de la supervivència és el mètode amb el qual és possible obtenir les corbes de supervivència^[42] a partir de les dades creades anteriorment. La fórmula és la següent:

$$\hat{S}(t_{(f)}) = \prod_{t_i < t} \frac{n_i - d_i}{n_i}$$

On $\hat{S}(t_{(f)})$ és l'estimador Kaplan-Meier, que indica la probabilitat d'un individu de viure més enllà d'un determinat temps t_i ; n_i el número de persones que resten al temps t_i i d_i el número d'esdeveniments en aquest temps. Així, l'estimador Kaplan-Meier és el càlcul de la probabilitat de supervivència per a cada temps de l'estudi. Amb R és possible obtenir l'estimació de Kaplan-Meier mitjançant les comandes *Surv()* i *survfit()*, totes dues pertanyents al paquet *survival*.

La comanda *Surv()* serveix per crear l'objecte que contindrà tots els temps de supervivència, diferenciant entre els individus que fallen i els censurats. Els dos objectes de supervivència, un per a cada tractament, s'han creat de la següent manera:

```
> superv.v <- Surv(dades2$temps,dades2$status==1) [1:336]
> superv.d <- Surv(dades2$temps,dades2$status==1) [337:672]
> superv.v[122:130]
[1] 7.75 7.88 9.25 9.25 10.25 0.00+ 0.06+ 0.12+ 0.19+
```

La variable *superv.v* inclou els temps de supervivència dels individus tractats amb vemurafenib i la variable *superv.d* el temps dels individus tractats amb dacarbazine. Com s'aprecia, la funció *Surv()* adjunta en alguns temps el símbol '+', indicant que aquests temps són de censura i diferenciant-los dels temps de falla, els quals es mostren sense cap afegit.

Amb els objectes de supervivència creats, la comanda *survfit()* permet el càlcul de l'estimador Kaplan-Meier, a més de donar dades com errors estàndard, intervals de confiança, medianes, etc. S'ha creat un objecte *survfit* per a cada tractament a partir dels objectes de supervivència realitzats:

```

> surv.v <- survfit(superv.v ~ 1, conf.type="plain")
> surv.d <- survfit(superv.d ~ 1, conf.type="plain")
> summary(surv.v)
Call: survfit(formula = superv.v ~ 1, conf.type = "plain")

   time n.risk n.event survival std.err lower 95% CI upper 95% CI
1.50   293     1    0.997 0.00341    0.9899    1.000
2.00   266     1    0.993 0.00505    0.9829    1.000
2.05   263     1    0.989 0.00629    0.9767    1.000
2.10   260     1    0.985 0.00732    0.9709    1.000
2.14   258     1    0.981 0.00823    0.9653    0.998
2.19   255     1    0.978 0.00905    0.9598    0.995
2.24   252     1    0.974 0.00981    0.9545    0.993
2.29   250     1    0.970 0.01052    0.9492    0.990
2.33   247     1    0.966 0.01118    0.9440    0.988

```

Es mostra un extracte de l'objecte *surv.v*, el qual conté els temps de falla (*time*), el número de persones en risc per a cada temps de falla (*n.risk*) i l'estimador Kaplan-Meier (*survival*), entre altres, dels individus tractats amb vemurafenib. Com s'observa, la probabilitat de supervivència va disminuint a mesura que passa el temps, fet lògic tenint en compte que, com més temps passa, més probabilitats hi ha que es produeixi la mort de l'individu. Dels objectes *surv.v* i *surv.d* encara se'n poden treure més dades d'interès, com per exemple la mediana:

```

> surv.v
Call: survfit(formula = superv.v ~ 1, conf.type = "plain")

      n  events  median 0.95LCL 0.95UCL
336.00 126.00   6.29   5.73   6.71
> surv.d
Call: survfit(formula = superv.d ~ 1, conf.type = "plain")

      n  events  median 0.95LCL 0.95UCL
336.00 163.00   4.36   3.46   5.09

```

La mediana indica la vida mitja de cada grup d'estudi, és a dir, el temps en què el número d'individus s'ha reduït a la meitat. Com s'observa, la mediana per al temps de supervivència és més alta en els individus tractats amb vemurafenib (*surv.v*) que amb els tractats amb dacarbazine (*surv.d*), essent de 6.29 mesos en el primer cas i 4.36 mesos en el segon. No obstant, en no treballar amb les dades adequades, aquests temps de vida mitja no s'han de veure com estimacions reals, però sí que serveixen per comprovar que el tractament amb vemurafenib provoca un augment del temps de vida mitja en els pacients amb melanoma. És moment de comprovar, doncs, si aquest augment s'allarga durant tot el temps d'estudi i si la diferència observada és significativa o no.

4.2.2.3. Estimació de les corbes de supervivència i càlcul del *Log-rank test*

Bàsicament, l'estimació de les corbes de supervivència consisteix en representar gràficament (Figura 4.14) les estimacions de Kaplan-Meier calculades a l'apartat anterior. Per tant, el procediment a seguir és semblant a l'anterior:

```
> color <- c("brown4", "cadetblue4")
> plot(survfit(Surv(dades2$temps, dades2$status)~dades2$tractament),
+      main="Estimació de la supervivència",
+      xlab="Temps de supervivència (mesos)",
+      ylab="Probabilitat de supervivència",
+      col=color)
> legend("bottomleft", c("Dacarbazine", "Vemurafenib"), lty=c(1,1),
+      col=color)
```

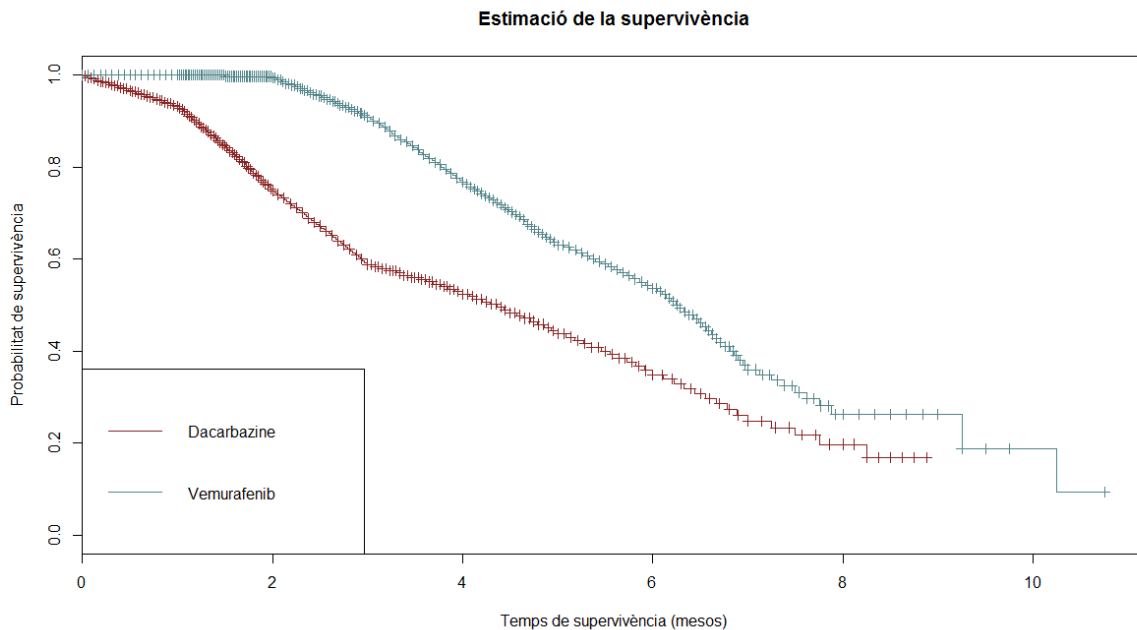


Figura 4.14. Gràfic d'estimació de les corbes de supervivència pels individus tractats amb vemurafenib (línia blava) i dacarbazine (línia grana). Les línies verticals de les corbes representen els individus censurats durant el temps d'estudi. A l'eix de les X es mostra el temps de supervivència (mesos) i a l'eix de les Y el percentatge de supervivència.

Com s'observa, el gràfic d'estimació de les corbes de supervivència té una estructura similar al gràfic del qual es basen les dades (Figura 4.13), essent l'estimació de la corba de supervivència major en els individus tractats amb vemurafenib que els tractats amb dacarbazine. No obstant, les corbes de supervivència, tot i semblants, són molt més acusades que les corbes de la Figura 4.13, el que significa que les medianes són també més baixes que les medianes reals.

L'últim pas és comprovar si les diferències que s'aprecien en el gràfic i en les medianes calculades anteriorment són significatives, fet que es pot dur a terme mitjançant el test *Log-rank*^[42]. El test *Log-rank* es una prova χ^2 que permet comparar corbes de

supervivència basant-se en valors observats i valors esperats. La fórmula pel càlcul del *Log-rank* es la següent:

$$\text{Log-rank test} = \frac{(O_i - E_i)^2}{\text{Var}(O_i - E_i)}$$

On O_i són els valors observats i E_i el valors esperats. Si el resultat del test *Log-rank* supera el valor teòric de les taules χ^2 es rebutja la hipòtesis nul·la d'igualtat en les corbes de supervivència. R permet el càlcul del test *Log-rank* de manera senzilla a través de la funció *survdiff()*, obtenint un p-valor que permetrà rebutjar o no la hipòtesis nul·la:

```
> survdiff(Surv(dades2$temps, dades2$status)~dades2$tractament)
```

```
Call:
```

```
survdiff(formula = Surv(dades2$temps, dades2$status) ~ dades2$tractament)
```

	N	Observed	Expected	(O-E) ² /E	(O-E) ² /V
dades2\$tractament=D	336	163	114	21.4	35.8
dades2\$tractament=V	336	126	175	13.9	35.8

```
Chisq= 35.8 on 1 degrees of freedom, p= 2.14e-09
```

Treballant amb una significació del 5%, el p-valor tan baix rebutja clarament la hipòtesis nul·la plantejada i, per tant, s'accepta que existeixen diferències significatives entre les dues corbes de supervivència.

5. Conclusions

5.2. Conclusions del treball

En aquest projecte s'ha demostrat que la inhibició de la mutació ^{V600E}BRAF mitjançant el fàrmac vemurafenib és un mètode eficaç per aturar la proliferació de cèl·lules canceroses en els melanomes. Per una banda, l'anàlisi transcripcional ha manifestat que l'ús del fàrmac en cèl·lules canceroses provoca una disminució significativa en l'expressió de gens que tenen relació amb la proliferació i creixement cel·lular, el que implica un fre en el creixement del càncer. A més, també s'ha observat una disminució de l'expressió en gens involucrats en processos de reparació, organització i empaquetament del DNA.

Les cèl·lules canceroses són molt més resistents que les cèl·lules no canceroses, capaces de sobreviure amb el medi mínim o en condicions adverses. És a dir, tot i que s'ha demostrat que el fàrmac vemurafenib atura el desenvolupament del càncer, això no té perquè significar l'eliminació de les cèl·lules malignes i, per tant, la reducció de la mida del tumor. No obstant això, el fet de disminuir també l'expressió de gens relacionats amb la reparació i organització del DNA pot significar, justament, que el fàrmac vemurafenib a més d'aturar la proliferació cel·lular també provoca l'eliminació d'aquestes. Tot i suportar condicions severes, excessives desregulacions que afecten en el DNA poden conduir a l'apoptosi cel·lular.

Curiosament, una de les vies metabòliques que s'ha trobat sobreexpressada amb l'ús del vemurafenib és la via MAPK, xocant de ple amb el fet que B-Raf es troba inhibida i que, per tant, aquesta via no s'hauria d'expressar. Tot i això, a l'inici del projecte s'ha explicat que existeix més d'una via MAPK, així que el fet d'inhibir la via MAPK/ERK podria comportar un augment de l'expressió de les altres vies MAPK.

Per altra banda, l'anàlisi de supervivència apunta en la mateixa direcció que l'anàlisi transcripcional. S'ha comprovat que el grup d'individus tractats amb vemurafenib tenen major temps de vida que els individus que segueixen el tractament clàssic amb dacarbazine. Tot i treballar amb dades aproximades i no les reals, el p-valor obtingut del test rebutja de manera clara la igualtat en les corbes de supervivència i, per tant, confirma l'augment de l'eficàcia en els tractaments de melanoma.

5.2. Assoliment dels objectius plantejats

Tots els objectius exposats en aquest projecte s'han complert de manera satisfactòria, però no així tots els objectius que s'havien plantejat en començar el treball. A l'inici també s'havia proposat analitzar la interacció entre $V600E$ BRAF i vemurafenib, és a dir, per quins residus es produeix la interacció, els tipus d'enllaços i quina és l'energia d'unió entre ells. Finalment, la falta d'informació trobada i la necessitat de cert programari per dur a terme aquest tipus d'anàlisi van fer eliminar aquest objectiu i continuar amb el treball. Tot i ser un tema interessant, la informació aportada per aquest apartat no és necessària per fer més entenedor aquest projecte.

5.3. Seguiment de la planificació i metodologia

La planificació s'ha seguit segons el que es va plantejar a l'inici del projecte tot i que, evidentment, durant el transcurs d'aquest hi ha hagut canvis. Per començar, el fet de no dur a terme l'objectiu comentat a l'apartat anterior ja significa un canvi a la planificació inicial feta. Aquest apartat es pretenia realitzar abans de començar amb l'anàlisi de dades de microarray, i és per aquest motiu que el diagrama de Gantt presenta dos dies buits abans de l'inici d'aquest.

A més, és complicat limitar la recerca bibliogràfica quan s'està buscant constantment informació relacionada entre ella. Per exemple, tot i que la recerca d'informació sobre la via MAPK/ERK es va limitar als quatre primers dies, és possible trobar nova informació (o més actualitzada) quan s'està cercant sobre la quinasa B-Raf i complementar així la primera part. Per tant, la planificació de la part bibliogràfica no s'ha de veure com quelcom tancat.

Un altre aspecte a comentar és que, com s'ha pogut comprovar, un anàlisi de dades de microarray és molt més complex que un anàlisi de supervivència (o, com a mínim, els que s'han dut en aquest projecte). No obstant, a l'inici del projecte es va planejar un temps similar per a realitzar-los, el que va comportar una major dedicació d'hores per dia en el primer anàlisi. Vist en perspectiva, hagués estat més adient dedicar més temps a l'anàlisi transcripcional i menys a l'anàlisi de supervivència.

En relació a la metodologia, sense dubte el fet de no disposar de les dades reals en el segon anàlisi ha condicionat i limitat molt el resultat d'aquest. Tot i que les corbes de

supervivència obtingudes són semblants a l'estudi del qual es basen les dades, no s'ha volgut aprofundir més ja que totes les dades que es poguessin obtenir (com en el cas de la mediana de supervivència) podrien ser força diferents a les dades reals.

5.4. Línies de treball futur

En haver utilitzat dades d'altres estudis, tot el que s'ha dut a terme en aquest projecte ja s'havia realitzat amb anterioritat i, evidentment, els estudis han seguit avançant. En aquest estudi s'ha comprovat que la inhibició de ^{V600E}BRAF mitjançant l'ús de vemurafenib frena la proliferació de cèl·lules canceroses, i actualment està demostrat que en dosis altes el fàrmac és capaç de provocar l'apoptosi cel·lular i disminuir la mida del tumor en un gran nombre de casos^{[28][30]}. A més, avui dia també existeix un altre fàrmac, anomenat dabrafenib, que és específic per combatre aquesta mutació^[29].

6. Referències

- 1 Kolch W (2000) Meaningful relationships: the regulation of the Ras/Ras/MEK/ERK pathway by protein interactions. *Biochem. J.* 351(2):289-305
- 2 Johnson GL, Lapadat R (2002) Mitogen-activated protein kinase pathways mediated by ERK, JNK, and p38 protein kinases. *Science* 298(5600):1911-1912
- 3 Mercer K, Giblett S, Oakden A, Brown J, Marais R, Pritchard C (2005) A-Raf and Raf-1 work together to influence transient ERK phosphorylation and G1/S cell cycle progression. *Oncogene* 24:5207-5217
- 4 Yamaguchi K, Shirakabe K, Shibuya H, Irie K, Oishi I, Ueno N, Taniguchi T, Nishida E, Matsumoto K (2008) Identification of a member of the MAPKKK family as a potential mediator of TGF-beta signal transduction. *Science* 270(5244):2008-2011
- 5 Roskoski R (2010) Raf protein-serine/threonine kinases: Structure and regulation. *Biochem. Biophys. Res. Commun.* 399(3):313-317
- 6 Zhang BH, Guan KL (2000) Activation of B-Raf kinase requires phosphorylation of the conserved residues Thr598 and Ser601. *EMBO J.* 19(20):5429-5439
- 7 Matallanas D, Birtwistle M, Romano D, Zebisch A, Rauch J, von Kriegsheim A, Kolch W (2011) Raf family kinases: old dogs have learned new tricks. *Genes cancer* 2(3):232-260
- 8 Tuveson DA, Weber BL, Herlyn M (2003) BRAF as a potential therapeutic target in melanoma and other malignancies. *Cancer Cell* 4(2):95-98
- 9 Leicht DT, Balan V, Kaplun A, Singh-Gupta V, Kaplun L, Dobson M, Tzivion G (2007) Raf kinases: Function, regulation and role in human cancer. *Biochim. Biophys. Acta* 1773(8):1196-1212
- 10 Madhunapantula SV, Robertson GP (2008) Is B-Raf a good therapeutic target for melanoma and other malignancies? *Cancer Res.* 68(1):5-8
- 11 Chong H, Lee J, Guan KL (2001) Positive and negative regulation of Raf kinase activity and function by phosphorylation. *EMBO J.* 20(4):3716-3727
- 12 Karasarides M, Chiloehes A, Hayward R, Niculescu-Duvaz D, Scanlon I, Friedlos F, Ogilvie L, Hedley D, Martin J, Marshall CJ, Springer CJ, Marais R (2004) B-RAF is a therapeutic target in melanoma. *Oncogene* 23(37):6292-6298
- 13 Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, Teague J, Woffendin H, Garnett MJ, Bottomley W, Davis N, Dicks E, Ewing R, Floyd Y, Gray K, Hall S, Hawes R, Hughes J, Kosmidou V, Menzies A, Mould C, Parker A, Stevens C, Watt S, Hooper S, Wilson R, Jayatilake H, Gusterson BA, Cooper C, Shipley J, Hargrave D, Pritchard-Jones K, Maitland N, Chenevix-Trench G, Riggins GJ, Bigner DD, Palmieri G, Cossu A, Flanagan A, Nicholson A, Ho JW, Leung SY, Yuen ST, Weber BL, Seigler HF, Darrow TL, Paterson H, Marais R, Marshall CJ, Wooster R, Stratton MR, Futreal PA (2002) Mutations of the BRAF gene in human cancer. *Nature* 417(6892):949-954
- 14 Hodis E, Watson IR, Kryukov GV, Arold ST, Imielinski M, Theurillat JP, Nickerson E, Auclair D, Li L, Place C, Dicara D, Ramos AH, Lawrence MS, Cibulskis K, Sivachenko A, Voet D, Saksena G, Stransky N, Onofrio RC, Winckler W, Ardlie K, Wagle N, Wargo J, Chong K, Morton DL, Stemke-Hale K, Chen G, Noble M, Meyerson M, Ladbury JE, Davies MA, Gershenwald JE, Wagner SN, Hoon

- DS, Schadendorf D, Lander ES, Gabriel SB, Getz G, Garraway LA, Chin L (2012) A landscape of driven mutations in melanoma. *Cell* 150(2):251-263
- 15 Roberts A, Allanson J, Jadico SK, Kavamura MI, Noonan J, Opitz JM, Young T, Neri G (2006) The cardiofaciocutaneous syndrome. *J. Med. Genet.* 43(11):833-842
 - 16 Sarkozy A, Carta C, Moretti S, Zampino G, Digilio MC, Pantaleoni F, Scioletti AP, Esposito G, Cordeddu V, Lepri F, Petrangeli V, Dentici ML, Mancini GM, Selicorni A, Rossi C, Mazzanti L, Marino B, Ferrero GB, Silengo MC, Memo L, Stanzial F, Faravelli F, Stuppia L, Puxeddu E, Gelb BD, Dallapiccola B, Tartaglia M (2009) Germline BRAF mutations in Noonan, LEOPARD, and cardiofaciocutaneous syndromes: Molecular diversity and associated phenotypic spectrum. *Hum. Mutat.* 30(4):695-702
 - 17 Li WQ, Kawakami K, Ruszkiewicz A, Bennett G, Moore J, Iacopetta B (2006) BRAF mutations are associated with distinctive clinical, pathological and molecular features of colorectal cancer independently of microsatellite instability status. *Mol. Cancer* 5:2
 - 18 Brose MS, Volpe P, Feldman M, Kumar M, Rishi I, Gerrero R, Einhorn E, Herlyn M, Minna J, Nicholson A, Roth JA, Albelda SM, Davies H, Cox C, Brignell G, Stephens P, Futreal PA, Wooster R, Stratton MR, Weber BL (2002) BRAF and RAS mutations in human lung cancer and melanoma. *Cancer Res.* 62(23):6997-7000
 - 19 Kimura ET, Nikiforova MN, Zhu Z, Knauf JA, Nikiforov YE, Fagin JA (2003) High prevalence of BRAF mutations in thyroid cancer: Genetic evidence for constitutive activation of the RET/PTC-RAS-BRAF signaling pathway in papillary thyroid carcinoma. *Cancer Res.* 63(7):1454-1457
 - 20 Agaimy A, Terracciano LM, Dirnhofer S, Tornillo L, Foerster A, Hartmann A, Bihl MP (2009) V600E BRAF mutations are alternative early molecular events in a subset of KIT/PDGFR α wild-type gastrointestinal stromal tumours. *J. Clin. Pathol.* 62(7):613-616
 - 21 Horbinski C (2012) Something old and something new about molecular diagnostics in glioma. *Surg. Path. Clin.* 5(4):919-939
 - 22 Kurman RJ, Shih IM (2011) Molecular pathogenesis and extraovarian origin of epithelial ovarian cancer: Shifting the paradigm. *Hum. Pathol.* 42(7):918-931
 - 23 Cohen Y, Xing M, Mambo E, Guo Z, Wu G, Trink B, Beller U, Westra WH, Ladenson PW, Sidransky D (2003) BRAF mutation in papillary thyroid carcinoma. *J. Natl. Cancer Inst.* 95(8):625-627
 - 24 Paik PK, Arcila ME, Fara M, Sima CS, Miller VA, Kris MG, Ladanyi M, Riely GJ (2011) Clinical characteristics of patients with lung adenocarcinomas harboring BRAF mutations. *J. Clin. Oncol.* 29(15):2046-2051
 - 25 Halaban R, Zhang W, Bacchiocchi A, Cheng E, Parisi F, Ariyan S, Krauthammer M, McCusker JP, Kluger Y, Sznol M (2010) PLX4032, a selective BRAF^{V600E} kinase inhibitor, activates the ERK pathway and enhances cell migration and proliferation of BRAF^{WT} melanoma cells. *Pigment Cell Melanoma Res.* 23(2):190-200
 - 26 Sala E, Mologni L, Truffa S, Gaetano C, Bollag E, Gambacorti-Passerini C (2008) BRAF silencing by short hairpin RNA or chemical blockade by PLX4032 leads to

- different responses in melanoma and thyroid carcinoma cells. *Mol. Cancer Res.* 6(5):751-759
- 27 Hoeflich KP, Gray DC, Eby MT, Tien JY, Wong L, Bower J, Gogineni A, Zha J, Cole MJ, Stern HM, Murray LJ, Davis DP, Seshagiri S (2006) Oncogenic BRAF is required for tumor growth and maintenance in melanoma models. *Cancer Res.* 66(2):999-1006
 - 28 Bollag G, Hirth P, Tsai J, Zhang J, Ibrahim PN, Cho H, Spevak W, Zhang C, Zhang Y, Habets G, Burton EA, Wong B, Tsang G, West BL, Powell B, Shellooe R, Marimuthu A, Nguyen H, Zhang KYJ, Artis DR, Schlessinger J, Su F, Higgins B, Iyer R, D'Andrea K, Koehler A, Stumm M, Lin PS, Lee RJ, Grippo J, Puzanov I, Kim KB, Ribas A, McArthur GA, Sosman JA, Chapman PB, Flaherty KT, Xu X, Nathanson KL, Nolop K (2010) Clinical efficacy of a RAF inhibitor needs broad target blockade in *BRAF*-mutant melanoma. *Nature* 467:596-599
 - 29 Gibney GT, Zager JS (2013) Clinical development of dabrafenib in BRAF mutant melanoma and other malignancies. *Expert Opin. Drug Metab. Toxicol.* 9(7):893-899
 - 30 Flaherty KT, Puzanov I, Kim KB, Ribas A, McArthur GA, Sosman JA, O'Dwyer PJ, Lee RJ, Grippo JF, Nolop K, Chapman PB (2013) Inhibition of mutated, activated BRAF in metastatic melanoma. *N. Engl. J. Med.* 363(9):809-819
 - 31 Parmenter TJ, Kleinschmidt M, Kinross KM, Bond ST, Li J, Kaadige MR, Rao A, Sheppard KE, Hugo W, Pupo GM, Pearson RB, McGee SL, Long GV, Scolyer RA, Rizos H, Lo RS, Cullinane C, Ayer DE, Ribas A, Johnstone RW, Hicks RJ, McArthur GA (2014) Response of BRAF-mutant melanoma to BRAF inhibition is mediated by a network of transcriptional regulators of glycolysis. *Cancer Discov.* 4(4):423-433
 - 32 Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D (2002) The human genome browser at UCSC. *Genome Res.* 12(6):996-1006.
 - 33 Lovly C, Pao W, Sosman J (2016) Molecular Profiling of Melanoma. *My Cancer Genome* (página web). Disponible a: <https://www.mycancergenome.org/content/disease/melanoma> Últim accés: 18 de maig de 2017
 - 34 Park T, Yi S-G, Kang S-H, Lee SY, Lee Y-S, Simon R (2003) Evaluation of normalization methods for microarray data. *BMC Bioinform.* 4(1):33.
 - 35 Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W and Smyth GK (2015) *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43(7): e47.
 - 36 Kim HJ, Bar-Sagi D (2004) Modulation of signalling by Sprouty: a developing story. *Nat. Rev. Mol. Cell Biol.* 5(6):441-450
 - 37 Bollag G, Tsai J, Zhang J, Zhang C, Ibrahim P, Nolop K, Hirth P (2012) Vemurafenib: the first drug approved for BRAF-mutant cancer. *Nat. Rev. Drug Discov.* 11(11): 873-896
 - 38 Lovly C, Pao W, Sosman J (2015) BRAF (V600K) mutation in melanoma. *My Cancer Genome* (página web). Disponible a: <https://www.mycancergenome.org/content/disease/melanoma/braf/116> Últim accés: 18 de maig de 2017
 - 39 Karoulia Z, Wu Y, Ahmed TA, Qisheng X, Bollard J, Krepler C, Wu X, Zhang C, Bollag G, Herlym M, Fagin JA, Lujambio A, Gavathiotis E, Poulikakos PI (2016) An

- integrated model of RAF inhibitor action predicts inhibitor activity against oncogenic BRAF signaling. *Cancer Cell* 30(3):485-498
- 40 Carvalho BS, Irizarry RA (2010) A framework for oligonucleotide microarray preprocessing. *Bioinformatics* 26(19):2363-2367
 - 41 Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J, Dummer R, Garbe C, Testori A, Maio J, Hogg D, Lorigan P, Lebbe C, Jouary T, Schadendorf D, Ribas A, O'Day SJ, Sosman JA, Kirkwood JM, Eggermont AMM, Dreno B, Nolop K, Li J, Nelson B, Hoy J, Lee RJ, Flaherty KT, McArthur GA (2011) Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N. Engl. J. Med.* 364(26):2507-2516
 - 42 Kleinbaum DG, Klein M (2012) *Survival Analysis: a self-learning text*. New York, USA: Springer
 - 43 Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhans M, Dreszer TR, Fujita PA, Guruvadoo L, Haeussler M, Harte RA, Heitner S, Hickey G, Hinrichs AS, Hubley R, Karolchik D, Learned K, Lee BT, Li CH, Miga KH, Nguyen N, Paten B, Raney BJ, Smit AF, Speir ML, Zweig AS, Haussler D, Kuhn RM, Kent WJ (2015) The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res.* 43:670-81
 - 44 Wan PT, Garnett MJ, Roe SM, Lee S, Niculescu-Duvaz D, Good VM, Jones CM, Marshall CJ, Springer CJ, Barford D, Marais R (2004) Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. *Cell* 116(6):855-867
 - 45 EMBL-EBI (2017) ChEMBL (pàgina web). Disponible a: <https://www.ebi.ac.uk/chembl>. Últim accés: 19 de maig de 2017
 - 46 NCBI (2017) GEO Accession viewer (página web). Disponible a: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE42872>. Últim accés: 19 de maig de 2017
 - 47 Sánchez A, Ruíz de Villa MC (2016) Análisis de datos de microarray. Universitat Oberta de Catalunya.
 - 48 Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry M, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene Ontology: tool for the unification of biology. *Nat. Genet.* 25(1):25-29
 - 49 The Gene Ontology Consortium (2015) Gene Ontology Consortium: going forward. *Nucl. Acids Res.* 43 Database issue D1049–D1056
 - 50 Kanehisa, Furumichi M, Tanabe M, Sato Y, Morishima K (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45:353-361
 - 51 Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44:457-462
 - 52 Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28:27-30
 - 53 Therneau TM, Grambsch PM (2000). *Modeling Survival Data: Extending the Cox Model*. Springer, New York. ISBN 0-387-98784-3.
 - 54 MacDonald JW (2017) hugene10sttranscriptcluster.db: Affymetrix hugene10 annotation data (chip hugene10sttranscriptcluster). R package version 8.6.0

- 55 Smith CA (2010) annaffy: Annotation tools for Affymetrix biological metadata. R package version 1.48.0
- 56 Falcon S, Gentleman R (2007). Using GOstats to test gene lists for GO term association. *Bioinformatics* 23(2):257-258
- 57 Carlson M (2017). org.Hs.eg.db: Genome wide annotation for Human. R package version 3.4.1
- 58 Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. *Nucleic Acids Res.* 28:235-242.
- 59 The UniProt Consortium (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 45:158-169

7. Annex

1. Seqüència nucleotídica de la regió codificant de BRAF, obtinguda a través del servidor web UCSC.

```
>hg38_knownGene_uc003vwc.5 range=chr7:140734479-140924764 5'pad=0 3'pad=0 strand=- repeatMasking=none
CGCCTCCCTTCCCCCTCCCCGCCGACAGCGGCCGCTCGGGCCCCGGCTC
TCGGTTATAAGATGGCGGCGCTGAGCGGTGGCGGTGGTGGCGGCGGGAG
CCGGGCCAGGCTCTGTTCAACGGGGACATGGAGCCCGAGGCCGGCGCCGG
CGCCGGCGCCGGGCTCTTCGGCTGCGGACCTGCCATTCCGGAGGAGG
TGTGGAATATCAAAACAAATGATTAAGTTGACACAGGAACATATAGAGGCC
CTATTGGACAAATTTGGTGGGGAGCATAATCCACCATCAATATATCTGGA
GGCCTATGAAGAATACACCAGCAAGCTAGATGCACTCCAACAAAGAGAAC
AACAGTTATTGGAATCTCTGGGGAACGGAACCTGATTTTTCTGTTTCTAGC
TCTGCATCAATGGATACCGTTACATCTTCTCCTCTTCTAGCCTTTCACT
GCTACCTTCATCTCTTTCAGTTTTTCAAATCCCACAGATGTGGCACGGA
GCAACCCCAAGTCAACCAAAAACCTATCGTTAGAGTCTTCTGCCCAAC
AAACAGAGGACAGTGGTACCTGCAAGGTGTGGAGTTACAGTCCGAGACAG
TCTAAAGAAAAGCACTGATGATGAGAGGTCTAATCCCAGAGTGTGTGCTG
TTTACAGAATTGAGGATGGAGAGAAAGAAACCAATTGGTTGGGACACTGAT
ATTTCTGGCTTACTGGAGAAGAATGTCATGTGGAAGTGTGGAGAAATGT
TCCACTTACAACACACAACCTTTGTACGAAAAACGTTTTTCACTTAGCAT
TTTGTGACTTTTGTGAAAAGCTGCTTTTCCAGGGTTTCCGCTGTCAAACA
TGTGGTTATAAAATTTCAACAGCGTTGTAGTACAGAAGTTCCACTGATGTG
TGTAAATTATGACCAACTTGATTTGCTGTTGTCTCCAAGTCTTTTGAAC
ACCACCCCAATACCACAGGAAGAGGGCTCCTTAGCAGAGACTGCCCTAACA
TCTGGATCATCCCCTTCCGCACCCGCTCGGACTCTATTGGGCCCAAAAT
TCTCACCAGTCCGCTCTCTTCAAATCCATTCCAATCCACAGCCCTTCC
GACCAGCAGATGAAGATCATCGAAATCAATTTGGCAACGAGACCAGTCC
TCATCAGCTCCCAATGTGCATATAAACACAATAGAACCCTGTCAATATTGA
TGACTTGATTAGAGACCAAGGATTTCTGTTGATGGAGGATCAACCACAG
GTTTGTCTGCTACCCCCCTGCCTCATTACCTGGCTCACTAACTAACGTG
AAAGCCTTACAGAAATCTCCAGGACCTCAGCGAGAAAGGAAAGTCACTTC
ATCCTCAGAAGACAGGAATCGAATGAAAACACTTGGTAGACGGGACTCGA
GTGATGATTGGGAGATTCTGATGGGCAGATTACAGTGGGACAAAGAATT
GGATCTGGATCATTTGGAACAGTCTACAAGGGAAAGTGGCATGGTGTGTG
GGCAGTGAATAATGTTGAATGTGACAGCACCTACACCTCAGCAGTTACAAG
CCTTCAAATGAAAGTAGGAGTACTCAGGAAAACACGACATGTGAATATC
TACTCTTATGGGCTATTCCACAAAGCCACAACCTGGCTATTGTTACCCA
GTGGTGTGAGGGCTCCAGCTTGTATCACCATCTCCATATCATTGAGACCA
AATTTGAGATGATCAAATTTATAGATATTGCACGACAGACTGCACAGGGC
ATGGATTAATACAGCCAAAGTCAATCATCCACAGAGACCTCAAGAGTAA
TAATATATTTCTCATGAAGACCTCAGTAAAAATAGGTGATTTGGTCT
TAGCTACAGTGAATCTCGATGGAGTGGTCCCATCAGTTTGAACAGTTG
TCTGGATCCATTTGTGGATGGCACCAGAAGTCAAGAAATGCAAGATAA
AAATCCATACAGCTTTCAGTCAAGTGTATATGCATTTGGAATTGTTCTGT
ATGAATGATGACTGGACAGTTACCTTATCAAACATCAACAACAGGGAC
CAGATAATTTTATGGTGGGACGAGGATACCTGTCTCCAGATCTCAGTAA
GGTACGGAGTAACTGTCAAAGCCATGAAGAGATTAATGGCAGAGTGCC
TCAAAAAGAAAAGAGATGAGAGACCACTTTTCCCAAAATCTCGCCTCT
ATTGAGCTGTGGCCCGCTCATTGCCAAAAATTCACCGCAGTGCATCAGA
ACCCTCCTGAATCGGGCTGGTTTCAAACAGAGGATTTAGTCTATATG
CTTGTGCTTCTCAAACACCCATCCAGGCAGGGGGATATGGTGCCTTT
CCTGTCCAAGTGAACAAATGAGTGAAGAGTTTCAAGAGAGTGAACAA
AGGAAAAATAAATGAACATATGTTGCTTATATGTTAAATGAATAAATA
CTCTTTTTTTTTAAGGTGAACCAAGAA
```