

APROXIMACIÓ COMPUTACIONAL ALS PROCESSOS DE SENESCÈNCIA EN CÀNCER DE PULMÓ

Georgina Fuentes Páez



Grau en Biotecnologia

Tutora: Malu Calle

Vic, Juny de 2020

Agraïments

M'agradaria mostrar el meu més sincer agraïment a totes les persones, companys, amics i professors que durant aquesta formació contínua m'han donat suport. En especial aquest últim, durant la realització d'aquest projecte perquè d'alguna manera han contribuït en la seva execució. L'encoratjament que m'han impregnat ha aconseguit que mantingués una actitud proactiva en tot moment.

En primer lloc, volia agrair a la Dra. Malu Calle per haver-me permès formar part del seu equip i ser la meva guia durant aquest projecte de computació. Sense la seva ajuda no hauria sigut possible la realització d'aquest. Volia donar-li les gràcies pels seus consells, la seva visió crítica, el seu rigor científic i la seva exigència i preocupació perquè tot sortís perfecte, cosa que m'ha ajudat no només durant l'execució del projecte, sinó també en altres àmbits. He tingut l'oportunitat de veure tant la seva part científic-tècnica com la humana, i m'agradaria que quedés constància que és una de les coses que més he valorat durant la realització d'aquest treball.

Agrair al Dr. Daniel Muñoz per brindar-me l'oportunitat de treballar al seu grup de recerca de la Universitat de Cambridge durant l'estiu, i per la confiança dipositada tant durant l'estança com durant el procés d'aquest projecte. Gràcies a ell, avui en dia puc dir que he viscut la millor experiència de la meva vida, tant en l'àmbit professional com en el personal.

Agrair a tots els professionals que formen part del grup del Dr. Daniel Muñoz, però sobretot, amb un èmfasi molt especial, m'agradaria fer una menció a l'estudiant de doctorat de la Universitat de Cambridge, Estela González, una persona que admiro molt en tots els sentits. L'Estela va ser la meva mentora durant l'estança i des de llavors, un exemple a seguir. Vull agrair tot el que vam estar compartint, la seva ajuda incessant i el seu ànim constant tant durant l'estada com durant la realització del projecte. La confiança que em va dipositar i que encara ara em continua brindant ha sigut un encoratjament per a encarar el meu futur en vers la recerca.

A la meva família, les persones més importants que em rodegen. Per haver fet tot el possible i més, privant-se de goigs perquè pogués cursar aquesta titulació. Agrair la confiança plena i cega, però sobretot per l'educació brindada. Gràcies als meus pares, per ser el motor que donen sentit a la majoria de decisions que prenc. Gràcies a la meva germana per la paciència i les intervencions durant la realització del projecte.

Per acabar, m'agradaria agrair a en Rubén la tolerància, la calma i la serenitat durant els moments complicats. Gràcies per acompanyar-me durant l'experiència i sobretot, gràcies per ser la persona que més confia en mi i la que sempre em fa créixer.

Resum

Entre els diferents tipus de càncers, el de pulmó és dels que provoca major mortalitat a escala mundial. Això es deu a diferents causes, però la primordial és la ineficiència en la seva detecció primerenca i el fracàs del tractament.

El meu projecte s'ha centrat en l'estudi de la senescència cel·lular en el context del càncer de pulmó. La senescència cel·lular és el procés pel qual les cèl·lules deixen de dividir-se. Ara bé, encara que no es divideixin, les cèl·lules es mantenen metabòlicament actives i usualment adopten un fenotip immunogènic que consisteix en la secreció d'agents proinflamatoris (conegut com a SASP, senescence-associated secretory phenotype, en anglès), senyals de supervivència i activitat β -galactosidasa. Això, a més de l'augment en l'expressió de proteïnes implicades en el cicle cel·lular com p16, p21 i p53, són biomarcadors de senescència cel·lular.

En càncer de pulmó existeixen evidències de la presència de cèl·lules senescentes en dos contextos diferents:

- (1) en lesions primàries precanceroses, on les cèl·lules del pulmó esdevenen senescentes en resposta a un estrès derivat de l'activació o sobreexpressió d'un oncogen, com ara K-Ras.
- (2) durant el tractament amb quimioteràpia en pacients amb tumors pulmonars avançats, on les cèl·lules tumorals esdevenen senescentes en resposta a un dany en l'ADN com a resultat de l'acció del fàrmac administrat (normalment, Cisplatí o Carboplatí en combinació amb Docetaxel o Taxol).

El principal problema amb la inducció de senescència és que hi ha evidències que quan aquestes cèl·lules s'acumulen poden tenir efectes secundaris adversos i perjudicials en el teixit. Per exemple, alguns estudis demostren que la presència de cèl·lules senescentes pot resultar en inflamació crònica, evasió del sistema immunitari i fins i tot, pot promoure la tumorigènesi a través dels factors que secreten a l'exterior (SASP).

En la meua estada de pràctiques externes al laboratori del Dr. Muñoz-Espín a la Universitat de Cambridge l'estiu de l'any 2019, vaig demostrar que quan les cèl·lules tumorals i fibroblàstiques de pulmó de ratolí són tractades amb Cisplatí, esdevenen en un tipus de senescència que a través del SASP, augmenta els nivells de proliferació, migració i creixement cel·lular tant de fibroblasts com de cèl·lules tumorals. Això pot indicar que el tractament amb els fàrmacs que s'acostumen a fer servir per a la quimioteràpia de càncer de pulmó, resulta en la inducció de senescència, denotant així que pot tenir efectes protumorigènics en altres cèl·lules, el que podria portar a una recaiguda o a la fallida del tractament.

Partint d'aquests resultats experimentals previs, en aquest treball ens plantegem l'estudi d'una signatura de senescència en càncer de pulmó i la seva implicació en diferents processos tumorals. D'una banda, s'han analitzat dades de pacients amb càncer de pulmó i s'ha estudiat l'associació dels marcadors de senescència amb variables clíniques com la supervivència, la quimioteràpia aplicada i l'estadi del tumor. Paral·lelament, amb mostres de ratolí, s'han comparat els perfils de senescència entre lesions primerenques o 'hiperplàstiques' amb lesions de tumors

avançats o ‘full-blown adenocarcinoma’ en càncer de pulmó de ratolí per determinar si les pautes d’expressió gènica dels tumors primerencs presenten una signatura de senescència cel·lular.

Per a dur a terme aquest treball es va recórrer a dades genòmiques recollides en diferents bases de dades públiques (*TCGA* i *GEO*), sobre les que es va realitzar una anàlisi d’aproximació computacional, mitjançant l’ús del llenguatge de programació *R*. Els Scripts pertinents contenint tots els codis emprats durant les anàlisis d’aquest estudi es van penjar a un repositori de GitHub, els podeu recórrer amb el següent enllaç: “https://github.com/georginafp/Senescence_lung_cancer”. En aquest hi trobareu un breu resum dels objectius del projecte i en una carpeta anomenada ‘Src’ hi trobareu els Scripts, els quals a més, estan implementats amb ‘hashtags’ amb el fi de facilitar l’enteniment de cada codi.

En dur a terme aquest estudi, s’han obtingut evidències que existeixen diferències significatives en l’àmbit d’expressió en molts processos biològics i de senyalització entre els grups analitzats. Finalment, es conclou que no hi ha evidència d’una signatura de senescència en pacients tractats amb quimioteràpia ni en lesions primerenques, rebutjant així la nostra hipòtesi experimental.

Índex

| | | |
|------------|---|----|
| 1. | Introducció | 1 |
| 1.1. | Punt de partida del projecte..... | 1 |
| 1.2. | Càncer | 1 |
| 1.2.1. | Principals tipus de càncers | 2 |
| 1.2.2. | Càncer de pulmó | 2 |
| 1.2.2.1. | Etiologia del càncer de pulmó | 2 |
| 1.2.2.2. | Tractament del càncer de pulmó..... | 3 |
| 1.2.2.2.1. | Cisplatí | 3 |
| 1.3. | Senescència | 3 |
| 1.3.1. | Marcadors de senescència cel·lular | 4 |
| 1.3.2. | SASP – Senescence Associated Secretory Phenotype..... | 4 |
| 1.3.2.1. | SASP i càncer | 4 |
| 1.3.3. | Senescència en càncer de pulmó..... | 5 |
| 1.4. | The Cancer Genome Atlas | 6 |
| 1.5. | Ciències òmiques | 6 |
| 1.5.1. | Transcriptòmica | 7 |
| 1.5.1.1. | Transcriptòmica i càncer | 7 |
| 1.6. | Tècniques d’anàlisi d’expressió diferencial..... | 8 |
| 1.6.1. | Tecnologia de RNA-sequencing | 8 |
| 1.6.2. | Tecnologia de Microarrays | 8 |
| 2. | Objectius | 10 |
| 3. | Metodologia | 12 |
| 3.1. | Dades públiques de càncer de pulmó..... | 12 |
| 3.2. | Estudi TCGA-LUAD | 12 |
| 3.3. | Estudi GSE65258..... | 16 |
| 4. | Resultats i discussió | 18 |
| 4.1. | Estudi TCGA-LUAD | 18 |
| 4.2. | Estudi GEO65258 | 35 |
| 5. | Conclusions..... | 40 |
| 6. | Bibliografia | 43 |

| | |
|---|----|
| Annexes..... | 45 |
| A.1. Evidències experimentals obtingudes durant l'estada de pràctiques a la Universitat de Cambridge..... | 45 |
| A.2. Evidències computacionals obtingudes durant el projecte d'anàlisi de dades de RNA-seq | 55 |

1. Introducció

1.1. Punt de partida del projecte

Aquest projecte es basa en les evidències que vaig obtenir durant l'estiu de l'any 2019 realitzant una estada de pràctiques externes a la Universitat de Cambridge. Les pràctiques les vaig fer al laboratori del Dr. Muñoz-Espín, que estudia el paper de la senescència cel·lular durant la iniciació i la progressió del càncer de pulmó. En els tres mesos d'estància vaig demostrar que quan les cèl·lules tumorals i fibroblàstiques de pulmó de ratolí són tractades amb Cisplatí, esdevenen en un tipus de senescència que a través del SASP, augmenta els nivells de proliferació, migració i creixement cel·lular tant de fibroblasts com de cèl·lules tumorals. Això pot indicar que el tractament amb els fàrmacs que s'acostuma a fer servir per a la quimioteràpia de càncer de pulmó, resulta en la inducció de senescència, denotant així que pot tenir efectes protumorigènics en altres cèl·lules, el que podria portar a una recaiguda o a la fallida del tractament. És per això que es proposa aquest estudi bioinformàtic, per tal de trobar una signatura de senescència en càncer de pulmó i determinar-ne així la seva implicació en diferents processos tumorals.

1.2. Càncer

Càncer és el nom que s'assigna a una col·lecció de malalties relacionades les quals s'inicien perquè alguna de les cèl·lules es comença a dividir sense detenció ni control fins a arribar a teixits circumdants, procés conegut com a metastasi (Corral, 2016).

En situacions fisiològiques, les cèl·lules creixen i es divideixen per a formar-ne de noves a mesura que el cos les necessita, i quan envelleixen, es danyen o moren, les noves són les que ocupen el seu lloc ("What Is Cancer? - National Cancer Institute", 2015). En canvi, en un procés de carcinogènesi, s'indueixen aberracions genètiques que transformen les cèl·lules normals en malignes (*figura 1*). És en aquest punt quan s'evadeixen dels senyals d'anticreixement i de diferenciació per tal d'adquirir un potencial de replicació il·limitat (Hanahan and Weinberg, 2011).

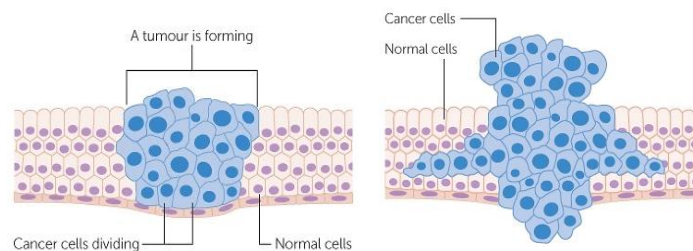


Figura 1. Iniciació d'un procés de carcinogènesi. Una cèl·lula creix anormalment i es divideix de manera descontrolada en un teixit determinat compost de cèl·lules normals. Canvis genètics com les mutacions provoquen que alguna d'aquestes cèl·lules es divideixi i prolifera de manera descontrolada formant l'estructura d'un tumor. Aquestes cèl·lules canceroses poden adquirir característiques de mesènquimes i així arribar a envair altres teixits. ("What Is Cancer? | Cancer Research UK" 2020., editat)

A més, doten de mecanismes diversos per tal d'afectar el microambient que les rodeja, promovent així processos com l'angiogènesi i l'evasió del sistema immunitari. És d'aquesta manera com aconsegueixen evitar els senyals d'apoptosis (Corral, 2016).

Aquestes alteracions genètiques o mutacions normalment són detectades durant el procés de divisió cel·lular i es poden reparar, però quan això no és possible o no es fa correctament i la cèl·lula defectuosa que s'havia de descartar no s'indueix apoptòsis, és quan es pot conduir a un procés de carcinogènesis (*Genes, ADN y càncer / Cancer Research UK, 2020*).

1.2.1. Principals tipus de càncers

El càncer s'agrupa segons el tipus de cèl·lula amb la qual s'inicia. N'hi ha cinc tipus principals (*Types of cancer / Cancer Research UK, 2020*):

- Carcinoma: càncer que comença a la pell o als teixits que recobreixen o cobreixen òrgans interns. Existeixen diferents subtipus, incloent-hi l'adenocarcinoma, que és d'especial interès per a aquest projecte.
- Sarcoma: càncer que comença als teixits connectius o de suport, com els ossos, cartílags, greixos, músculs o vasos sanguinis.
- Leucèmia: càncer que comença al teixit formador de la sang com la medul·la òssia i que fa que es produeixin cèl·lules sanguínies anormals que ingressen a la sang.
- Limfoma i mieloma: càncers que comencen a les cèl·lules del sistema immune.
- Càncers de cervell i medul·la espinal: aquests dos tipus de càncers se'ls coneix com a càncers del sistema nerviós central.

1.2.2. Càncer de pulmó

El càncer de pulmó és una col·lecció heterogènia de tumors originats a l'epiteli de recobriment o glandular dels bronquis (Corral, 2016).

A les etapes primerenques del càncer de pulmó no hi ha símptomes evidents, però quan aquest avança els pacients desenvolupen una sèrie d'indicadors de la malaltia (*Lung cancer - NHS, 2020*). Entre ells el més característic és la tos, ja que l'origen del càncer de pulmó és a les vies aèries principals, i quan hi ha una dilatació dels ganglis limfàtics acompanyats d'una pneumonitis obstructiva, com a resultat es produeix la tos (Corral, 2016). A més, també hi podríem trobar sagnat i dolor al tossir, dispnea acompanyada de dolor, un cansament inexplicable i una pèrdua de pes important (*Lung cancer - NHS, 2020*).

1.2.2.1. *Etiologia del càncer de pulmó*

El tabaquisme és el factor més gran de risc relacionat amb el càncer de pulmó, però s'ha reportat que el risc i la susceptibilitat a agents carcinògens respiratoris depèn de la interacció entre l'exposició al carcinogen, en aquest cas el factor de risc, i la susceptibilitat individual al determinat carcinogen. Essent doncs, una interacció sinèrgica entre diferents factors de risc que fan de vital transcendència el risc total de patir càncer de pulmó (Corral, 2016).

Així i tot, els agents carcinògens expulsats en l'exhalació del fum, també en són la primera prevenció perquè les persones fumadores o les que estan en continu contacte amb els carcinògens del seu fum, tenen fins a vint vegades més risc de patir càncer de pulmó (Doll *et al.*, 2004).

Hi ha un component genètic relacionat amb el càncer de pulmó, i aquest pot provenir per l'historial de tabaquisme, l'exposició al fum del tabac o per antecedents familiars de càncer, i tots

ells augmenten de manera significativa la incidència a patir-lo (Dela Cruz, Tanoue and Matthay, 2011).

1.2.2.2. *Tractament del càncer de pulmó*

La teràpia que s'aplica en pacients amb càncer de pulmó depèn d'una varietat de factors com en són: el lloc i la mida del càncer, l'estadi en què es troba el tumor i la salut del pacient en general (*Lung cancer - Treatment - NHS*, 2020). Les opcions de tractament més comunes o convencionals es regeixen per la resecció quirúrgica, la quimioteràpia basada en platí i la radioteràpia. Essent aquestes últimes emprades tant per separat com en combinació (Corral, 2016).

En termes de supervivència global, control de símptomes i qualitat de vida, el tractament citotòxic ha demostrat beneficis estadísticament significatius. En canvi, els tractaments de quimioteràpia basats en platí, específicament Cisplatí, provoquen un augment d'entre un 20 i un 35% en la progressió tumoral (Corral, 2016).

1.2.2.2.1. Cisplatí

El cisplatí (cis-diamminedichloroplatinato (II)) és un fàrmac quimioterapèutic que s'ha inclòs en la teràpia de nombrosos tipus de tumors sòlids (Dasari and Bernard Tchounwou, 2014). Oimés, és un dels components principals en quimioteràpies curatives i pal·liatives (Cotarelo Fernández, 2016; Guancial *et al.*, 2016).

El seu mecanisme d'acció està basat en platí, el qual penetra les cèl·lules interferint en la divisió cel·lular (Cotarelo Fernández, 2016) i impeding els processos de transcripció i replicació amb l'objectiu d'induir aquella cèl·lula a apoptosi (Dasari and Bernard Tchounwou, 2014; Guancial *et al.*, 2016).

1.3. Senescència

La senescència cel·lular és el procés pel qual la cèl·lula envelleix i deixa de dividir-se de manera permanent però sense destruir-se. Amb el temps, grans quantitats d'aquest tipus de cèl·lules s'acumulen en els teixits de manera que si són actives, alliberen factors perjudicials que causen una inflamació que afecta negativament les cèl·lules veïnes (Collado and Serrano, 2010).

Aquest procés es dona un cop les cèl·lules han proliferat en excés o sota estímuls anormals potents, i la seva inducció podria ser causada per diverses situacions d'estrès com en són: el dany cel·lular oxidatiu, l'escurçament dels telòmers, danys en l'ADN i els fàrmacs de quimioteràpia com el Cisplatí (Collado and Serrano, 2010), essent aquest últim de gran importància per a aquest projecte.

Per esguard d'això, hi ha evidències que suggereixen que la senescència va emergir per tal de suprimir la tumorigènesi, actuant com un mecanisme de seguretat per prevenir la proliferació de cèl·lules amb risc de patir transformacions neoplàsiques (Collado and Serrano, 2010). En canvi, es creu que poden implementar un fenotip secretori proinflamatori i protumorigen en forma de complex anomenat SASP (*senescence associated secretory phenotype* – en anglès), el qual

afectaria les cèl·lules veïnes o al microambient accelerant el creixement del tumor (*figura 2*) (Muñoz-Espín *et al.*, 2013).

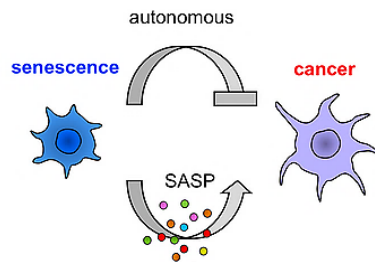


Figura 2. Efecte del fenotip secretori proinflamatori i protumorigen SASP. Aquest promou el desenvolupament i la progressió del càncer afectant el microambient i a les cèl·lules veïnes (*Identification of candidate processes and mechanisms involved in lung*, 2020).

1.3.1. Marcadors de senescència cel·lular

En primer lloc, les cèl·lules senescentes *in vitro* es caracteritzen per dotar d'una morfologia aplanada i ampla (*imatge 1 de l'annex*), però *in vivo* presenten la mateixa morfologia que l'etapa del tumor corresponent.

Malgrat els estudis i l'interès per a definir la senescència cel·lular, encara hi ha una manca de marcadors robustos per a la determinació d'aquest procés. El més acceptat i àmpliament utilitzat és la tinció de β -Galactosidasa (*imatges 1 i 2 de l'annex*) (avaluada a un pH òptim de 6), el qual és un enzim present a les cèl·lules senescentes, de manera que la seva detecció confirma l'estat de senescència (Braig *et al.*, 2005; Campo-Trapero *et al.*, 2008).

Altres marcadors clàssics de la senescència són les mateixes proteïnes implicades en el mecanisme d'aturada del creixement cel·lular, com ara els gens p14, p15, p16, Dec1, i Dcr2. Igual que la Ciclina D, p53, pRb, i p21 (Campisi, 2005; Campo-Trapero *et al.*, 2008).

1.3.2. SASP – Senescence Associated Secretory Phenotype

El SASP és una secreció d'entre 40 i 80 factors involucrats en diferents vies de senyalització intracel·lular. Aquests es poden dividir en dues categories principals (Yazmín González-Puertos *et al.*, 2015):

- (1) factors de senyalització solubles: com en serien clars exemples les interleucines (*IL1*, *IL6* i *IL8*), quimosines (cofactors de membrana (*MCO1*, *MCO2* i *MCO4*), *eotaxina* i proteïnes inflammatòries de macròfags (*MIP*)), factors de creixement (*IGF* i *CSF*) i metal·loproteases (*MMP*)
- (2) factors insolubles: compost per espècies reactives de l'oxigen (ROS), a més d'alguns ions i microRNAs.

1.3.2.1. *SASP i càncer*

Els factors que componen el SASP confereixen a les cèl·lules senescentes la característica de modificar el microambient extracel·lular, de manera que desencadenen un conjunt de vies de

senyalització que en modificaran les respostes a les cèl·lules adjacents. La seva anàlisi es complica perquè té un paper d'ambigüitat. Alguns factors afavoreixen l'efecte del deteriorament i el dany, mentre que altres en podrien ser beneficiosos (*figura 9*) (Yazmín González-Puertos *et al.*, 2015)

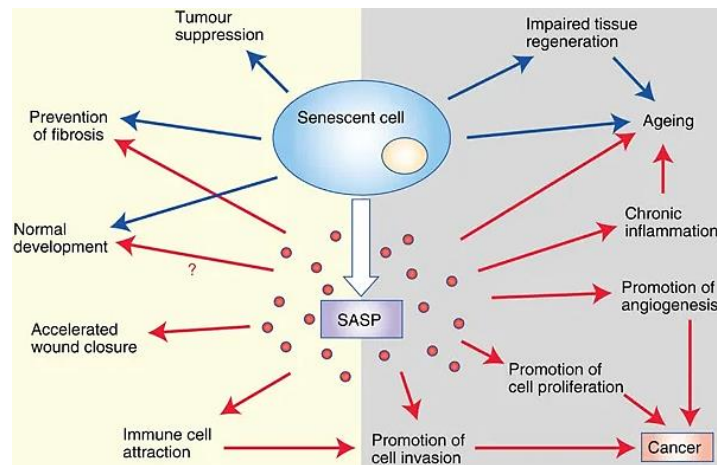


Figura 3. Dualitat present en la senescència cel·lular. Per una banda, al costat nítid hi trobem els efectes beneficiosos on es representa que l'arrest cel·lular evita la tumorigènesi, que evadint la proliferació de cèl·lules que segreguen ECM s'evita la fibrosi, que algunes característiques del desenvolupament embrionari normal són promogudes per cèl·lules senescentes, i finalment, que els components del SASP acceleren el tancament de les ferides i atrauen cèl·lules del sistema immunitari. Per una altra banda, al costat fosc hi trobem els efectes perjudicials on es representa que la pèrdua del potencial de proliferació que acompanyava a la senescència té un efecte en la regeneració de teixits afavorint l'envelliment, procés que es veu implementat per l'efecte del SASP, ja que provoca un estat inflamatori crònic en el microambient tissular. Finalment, s'assenyala com aquest secretoma conté factors que promouen l'angiogènesi, la proliferació cel·lular i el tret d'invasió agressiu de cèl·lules canceroses. Aquestes activitats finalment generen una progressió del càncer (Lecot *et al.*, 2016, editat).

Una funció important de SASP és l'estimulació del sistema immunitari per eliminar les cèl·lules senescentes i evitar-ne així la seva acumulació. A més, el SASP pot evitar el desenvolupament del càncer, ja que s'ha reportat que secreta factors antiangiogènics com la *maspina*. A més, podria induir l'apoptosi d'algunes línies cel·lulars, evitant així la proliferació de cèl·lules danyades (Yazmín González-Puertos *et al.*, 2015).

Al contrari, una gran quantitat de components de SASP promouen la inflamació crònica, la qual és una de les causes associades a pràcticament totes les malalties relacionades amb l'edat, a més de jugar un paper molt important en etapes primerenques del càncer. Oimés, aquest secretoma pot estimular la proliferació i diferenciació cel·lular així com l'angiogènesi per tal de formar nous vasos sanguinis. I fins i tot, impulsar la transició d'epiteli-mesènquima (TEM) afavorint la metastasi en les cèl·lules premalignes (Yazmín González-Puertos *et al.*, 2015).

1.3.3. Senescència en càncer de pulmó

Així doncs, veiem que la senescència cel·lular pot tenir efectes contradictoris en càncer. Avui dia, encara no s'entén del tot què determina que una cèl·lula senescent sigui beneficiosa i pugui aturar la progressió d'un tumor o bé contribueixi al fet que el tumor continuï creixent. En aquest sentit, entendre què succeeix durant el tractament amb quimioteràpia en pacients amb càncer de pulmó pot ser molt important.

Un estudi l'any 2005 va demostrar per primera vegada que la quimioteràpia estàndard basada en platí resulta en un augment significatiu de cèl·lules positives en β -Galactosidasa, el que pot indicar que la quimioteràpia indueix senescència als tumors pulmonars (Roberson *et al.*, 2005). Més endavant, un segon estudi l'any 2013 va demostrar que els pacients de càncer de pulmó que rebien quimioteràpia i presentaven més cèl·lules positives en β -Galactosidasa als tumors, tenien una pitjor supervivència que els que no presentaven aquest marcador als seus tumors (Wang *et al.*, 2013). No obstant això, els investigadors es van basar només en l'anàlisi d'aquesta única tinció i una cohort de només 18 pacients en total.

Tot i que aquests estudis poden indicar que la quimioteràpia en càncer de pulmó provoca senescència i això resulta en una pitjor prognosi per als pacients, per arribar a unes conclusions més evidents, es requereix una anàlisi més completa i amb un nombre de pacients més elevat. En aquest sentit, la bioinformàtica i l'ús de bases de dades poden ser unes eines molt importants. Aquestes bases recullen dades clíniques rellevants com el tractament i la supervivència de centenars de pacients, així com els perfils d'expressió gènica. Entre d'altres, que ens poden permetre determinar la presència de senescència en les mostres i correlacionar-les amb la supervivència.

1.4. The Cancer Genome Atlas

L'Institut Nacional del Càncer (NCI) junt amb l'Institut Nacional de la Investigació del Genoma Humà (NHGRI) van col·laborar per establir l'Atlas del Genoma del Càncer (TCGA) (<https://www.cancer.gov/>) amb la finalitat de caracteritzar tots els canvis genètics presents en els diferents tipus i subtipus de càncers (M. Parra, 2018).

Aquesta plataforma ha estat dissenyada per tal que totes aquestes col·leccions de dades que fins ara han esdevingut una millora en termes de control de diagnòstic, de tractament i de prevenció pel càncer, a hores d'ara estiguin disponibles i a l'abast de qualsevol comunitat que les utilitzi (M. Parra, 2018).

1.5. Ciències òmiques

Les ciències òmiques és un conjunt de ciències que es caracteritzen per l'addició d'un sufix "-òmica-" per a referir-se a "conjunt de". Aquest conjunt de ciències engloba la genòmica que és l'estudi de la seqüència genètica en la seva totalitat, la transcriptòmica que es regeix per l'estudi del conjunt d'ARNm transcrits a la cèl·lula, la metabolòmica que es basa la quantificació i l'estudi del conjunt de metabòlits presents a la cèl·lula en un estat fisiològic i la proteòmica que és l'estudi del conjunt proteic en un moment i teixit determinats (Marín, 2014).

Les tecnologies d'alt rendiment són les que s'usen per a l'estudi de les òmiques, i junt amb la seva instrumentària generen grans quantitats de dades d'un volum molt extens, que avui dia encara segueix creixent de manera accelerada. És per això que la integració d'aquesta gran quantitat de dades per tal de donar resposta a moltes preguntes no encara resoltes, suposa tot un repte per la ciència (Marín, 2014).

En definitiva, tots els avenços tecnològics combinats amb la bioinformàtica permeten abastar les qüestions biològiques gràcies a la combinatòria de les ciències òmiques (Casado-Vela *et al.*, 2011).

1.5.1. Transcriptòmica

Es descriu la transcriptòmica com l'estudi del transcriptoma, essent aquest el conjunt complet de transcrits d'ARN que produeix el genoma tant en circumstàncies específiques com en una cèl·lula específica (*Transcriptomics - Latest research and news | Nature*, 2020).

L'expressió gènica és la manifestació fenotípica d'un gen o gens per mitjà de processos com la transcripció i la traducció genètica (*figura 4*). De manera que, les anàlisis d'expressió gènica determinen el patró de gens expressats sota circumstàncies o cèl·lules específiques (NCBI - National Center For Biotechnology Information, 2020).

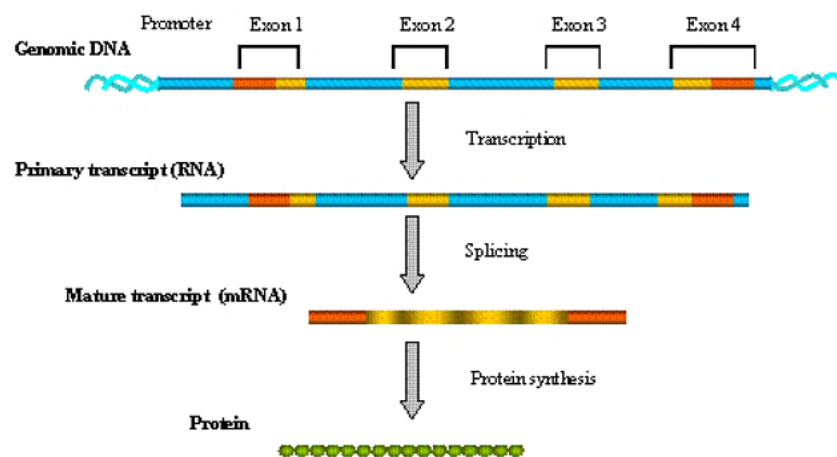


Figura 4. Dogma central de la biologia molecular. La informació genètica que es troba en seqüències de bases nitrogenades en l'ADN es transcriu a una molècula d'ARN. Aquestes molècules surten del nucli cel·lular per tal d'arribar al citoplasma, ja que aquest és l'espai on es dona la síntesi proteica, procés conegut com a traducció (NCBI - National Center For Biotechnology Information, 2020).

La comparació de transcriptomes doncs, permet identificar gens que s'expressen diferencialment en diferents poblacions cel·lulars, o en resposta a diferents tractaments (*Transcriptomics - Latest research and news | Nature*, 2020).

1.5.1.1. *Transcriptòmica i càncer*

La disponibilitat de dades de perfils d'expressió de múltiples tumors i les tècniques de seqüenciació massiva han promogut evidències sobre la importància que té el transcriptoma en el desenvolupament de càncers (M. Parra, 2018), ja que facilita el diagnòstic i augmenta la precisió en la predicció dels pronòstics (Sager *et al.*, 2015).

Encara que el càncer s'origini per mutacions en la seqüència d'ADN, aquestes mutacions afecten el transcriptoma i poden induir mecanismes que estiguin vinculats amb el desenvolupament del càncer (Sebestyén *et al.*, 2016). Per tant, no només té importància la mutació

puntual que ha fet emergir el càncer, sinó que també ho és el coneixement de si aquell gen s'expressa i com ho fa (M. Parra, 2018).

1.6. Tècniques d'anàlisi d'expressió diferencial

D'entre els mètodes per a mesurar els nivells d'expressió gènica, per a aquest projecte s'han utilitzat dades generades a partir de tant la tècnica de seqüenciació massiva coneguda com a *RNA-sequencing* (RNA-seq) com a partir de la tecnologia de Microarrays.

1.6.1. Tecnologia de RNA-sequencing

El primer pas és l'extracció d'ARN missatger (ARNm) madur, el qual conté només la seqüència codificant. Posteriorment es fragmenta i mitjançant una transcripció inversa, aquest es converteix en ADN complementari (ADNc) el qual després se seqüenciarà. Es procedeix amb l'alineament de les seqüències obtingudes en vers unes seqüències d'un genoma de referència, reconstruint així, totes les regions del genoma que es transcriuen (M. Parra, 2018) com es pot observar a la *figura 5*.

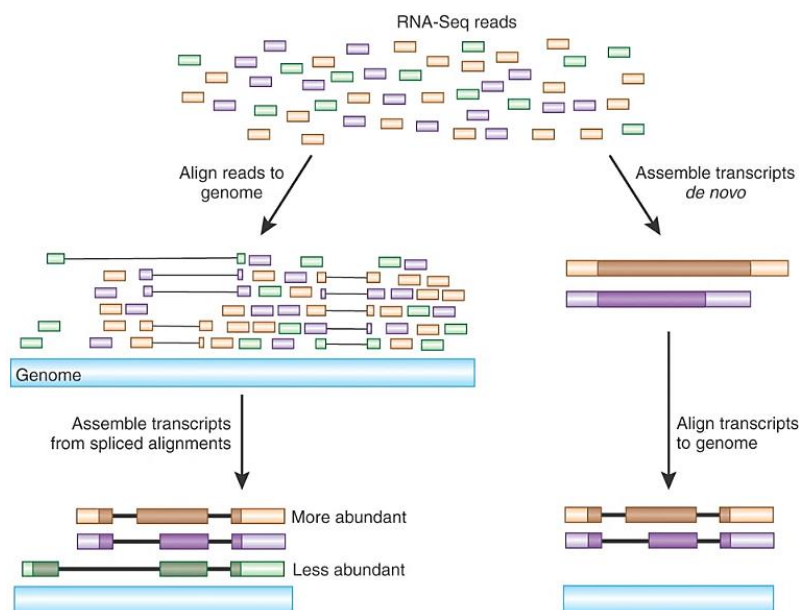


Figura 5. Comparativa entre els dos mètodes d'alineament i acoblament de la tècnica d'RNA-seq. A l'esquerra es troba representat el primer mètode, on s'alineen els "reads" de curta llargada amb el genoma de referència tenint en compte els possibles "overlappings", i paral·lelament es procedeix a la reconstrucció dels transcrits. En canvi, a la dreta hi trobem representat el segon mètode, on primer s'acoblen les seqüències de transcripció "de novo", és a dir, directament de les lectures d'RNA-seq, i s'alinea amb el genoma de referència. S'hi diferencien tonalitats de color segons la isoforma de transcripció de la qual es tracti (Haas and Zody, 2010).

1.6.2. Tecnologia de Microarrays

La tecnologia de microarrays és una tècnica que estudia l'expressió de múltiples gens alhora. Consisteix a col·locar milers de seqüències gèniques en llocs determinats d'una plataforma de vidre, coneguda com a 'chip', i mitjançant un aparellament per complementarietat entre les bases de la mostra i les seqüències de les sondes presents en el chip, es produeix una quantitat de llum mesurable. Les àrees del chip que produeixen llum identifiquen els gens que s'expressen en

aquella mostra, indicant doncs, el nivell d'expressió (Casellas. A, UVic, Assignatura de Genòmica, 2020).

Les aplicacions d'aquesta tecnologia entre d'altres inclouen l'estudi de gens que s'expressen diferencialment entre diverses condicions que és la de major interès per a aquest projecte.

2. Objectius

L'objectiu principal d'aquest treball és l'estudi computacional dels biomarcadors de senescència en càncer de pulmó i la seva implicació en diferents processos tumorals.

D'una banda es pretén analitzar dades de pacients amb càncer de pulmó i estudiar l'associació dels marcadors de senescència amb variables clíniques com la supervivència, la quimioteràpia aplicada i l'estadi del tumor. L'objectiu és demostrar la hipòtesi següent: els pacients que reben quimioteràpia tenen major expressió de marcadors de senescència en els seus tumors i, per conseqüència, una pitjor prognosi (taxa de supervivència disminuïda).

Per això cal fer una recerca de bases de dades que recullin dades d'expressió gènica en tumors de pacients de càncer de pulmó, i que permetin discernir entre pacients que hagin rebut quimioteràpia o no abans d'extreure i analitzar l'expressió gènica en el tumor. Els objectius específics d'aquest estudi inclourien respondre a les següents preguntes:

- (1) Analitzar l'efecte de la quimioteràpia sobre la supervivència dels pacients amb càncer de pulmó:
 - a. Quina és la diferència en la taxa de supervivència en pacients que han rebut quimioteràpia en comparació amb aquells que no han estat tractats (sobreviuen durant més temps, menys o igual)?
- (2) Analitzar les pautes d'expressió gènica dels tumors per veure si en els tumors de pacients que han rebut quimioteràpia trobem una signatura de senescència cel·lular.
 - a. Els tumors pretractats amb quimioteràpia, expressen marcadors de senescència com p16, p21, p53, IL-1, IL-6, IL-8, etc.?
 - b. És aquesta expressió significativament superior en els tumors pretractats, comparat amb els tumors que no han rebut quimioteràpia?
- (3) Finalment, trobar una relació entre la possible signatura de senescència en tumors i la supervivència dels pacients.
 - a. Quina és la taxa de supervivència dels pacients que tenen major expressió de marcadors de senescència (p16, p21, etc.)?
 - b. Si dividim als pacients en dos grups (alta vs. baixa expressió de marcadors de senescència), quin efecte trobem en la supervivència?

D'altra banda, paral·lelament es vol fer una anàlisi sobre la relació entre la senescència i el desenvolupament del càncer de pulmó a diferents estadis. Com hem dit, la senescència també es pot induir, no només a través de tractaments de quimioteràpia, sinó també per la sobreexpressió d'oncogens.

En aquest sentit, l'objectiu és comparar les lesions primerenques o 'hiperplàstiques' amb lesions de tumors avançats o 'full-blown adenocarcinoma' per tal de determinar si hi ha evidències de senescència en les lesions primerenques de càncer de pulmó de ratolí.

Analitzant doncs, les pautes d'expressió gènica dels tumors primerencs i veure si es troba una signatura de senescència cel·lular. Els objectius d'aquest estudi paral·lel inclourien respondre a les següents preguntes:

- a. Les lesions primerenques expressen marcadors de senescència com p16, p21, p53, IL-1, IL-6, IL-8, etc.?
- b. És aquesta expressió significativament superior en els tumors primerencs, comparat amb els tumors d'estadi avançat?

3. Metodologia

3.1. Dades públiques de càncer de pulmó

Per aquest estudi s'han utilitzat dos conjunts de dades públiques sobre càncer de pulmó. La primera cohort de pacients, extreta del repositori de l'Atlas del Genoma del Càncer (TCGA – *The Cancer Genome Atlas*, en anglès), conté dades d'expressió gènica obtingudes a partir de seqüenciació (RNA-seq). El segon estudi, extret de la plataforma Omnibus de l'expressió gènica (GEO – *Gene expression Omnibus* – en anglès), conté dades d'expressió gènica de ratolins obtingudes a partir de microarrays. Ambdós tipus de dades (seqüenciació o microarrays) requereixen un tractament bioinformàtic i estadístic específic que es descriu a continuació.

3.2. Estudi TCGA-LUAD

3.2.1. Cohort de pacients

El present estudi es va dur a terme mitjançant les dades del projecte *TCGA-LUAD* del repositori de l'Atlas del Genoma del Càncer (TCGA- *The Cancer Genome Atlas* – en anglès). Aquest repositori és la base més gran de dades d'informació genètica de càncer.

El projecte *TCGA-LUAD* conté dades de 522 pacients diagnosticats amb càncer de pulmó, incloent informació tant de perfil transcriptòmic (RNA-seq), com de dades clíniques (estadi del tumor, dies de seguiment, any en què es va fer el diagnòstic, cigarrets al dia, diagnòstic primari, edat, sexe, estat vital i tractament aplicat) (*Taula 1*). Els dos grups de pacients que ens interessa comparar són els pacients als quals no se'ls ha subministrat cap mena de tractament i aquells als que se'ls va subministrar un tractament farmacològic basat en platí, havent estat ambdós grups sota seguiment o havent arribat a la seva mort, segons sigui el cas, en un període llarg de temps.

Taula 1. Taula descriptiva de la informació continguda en el projecte TCGA-LUAD. La taula conté informació respecte a el nombre de pacients totals de l'estudi, el nombre de pacients tractats i no tractats, el sexe, l'estat vital, l'estadi tumoral i la franja d'edat dels pacients.

| Núm. pacients totals | Núm. pacients en vers el tractament | Núm. pacients en vers el sexe | Núm. pacients en vers l'estat vital | Núm. pacients en vers l'estadi tumoral | Núm. pacients segons l'edat |
|----------------------|---|-------------------------------|-------------------------------------|---|--|
| 522 | Quimioteràpia: 109 No quimioteràpia: 367 No reportats: 46 | Dones: 280 Homes: 242 | Vius: 334 Morts: 188 | Estadi I: 279 Estadi II: 124 Estadi III: 85 Estadi IV: 26 No reportats: 8 | Dels 30 als 50: 38 Dels 50 als 70: 297 Dels 70 als 90: 187 |

3.2.3. Descàrrega de dades d'accés obert

Es va treballar amb el paquet *TCGAbiolinks* per a poder descarregar les dades des del mateix repositori amb la possibilitat d'analitzar-les directament amb l'ús d'un sol paquet d'*R*. Per a l'obtenció de les dades de perfil transcriptòmic es va emprar la funció *TCGAquery*, mentre que per a l'obtenció de les dades de perfil clínic es va emprar la funció *TCGAquery_clinic*. Finalment, amb la funció *TCGAdownload* es van descarregar.

3.2.4. Preparació de les dades

La lectura i descàrrega de dades es va realitzar amb la funció *TCGAprepare*. Aquesta converteix les dades en un *summarizedExperiment*, el qual serà un objecte per a l'anàlisi *downstream*. Finalment, es guarda aquest objecte en un fitxer en format *.rda*.

3.2.5. Preprocessament de les dades

La funció *TCGAanalyze_Preprocessing* defineix una matriu simètrica quadrada de correlació de *Pearson* i un diagrama de caixes de correlació entre les mostres, la qual cosa permet trobar baixa correlació entre aquestes, definint quines podrien ser considerades com a *outliers*.

3.2.6. Normalització de les dades

La normalització permet comparar els nivells d'expressió entre diferents mostres i rèpliques (Marioni et al., 2008), ja que així assegurem que les diferències en el nombre de *reads* obtinguts reflecteixin l'expressió diferencial existent i no biaixos artificials de la seqüenciació. En aquest projecte es va emprar la funció *TCGAanalyze_Normalization* a partir del mètode de contingut en *GCs*.

3.2.7. Filtratge de les dades

Amb l'objectiu d'obtenir resultats significatius es van filtrar els transcrits seleccionant un llinar. En aquest projecte s'utilitza la funció *TCGAanalyze_Filtering* a partir del mètode per *quantils* i amb el llinar assignat a 0,25.

3.2.9. Anàlisi d'expressió diferencial

La determinació dels gens diferencialment expressats entre els grups d'interès (tractats i no tractats), es va realitzar mitjançant la funció *TCGAanalyze_DEA*. Per a representar els nivells d'expressió dels gens es va generar un *volcanoPlot* amb les funcions *with* i *subset*, i un *HeatMap* amb la funció *heatmap.2*.

3.2.10. Anàlisi d'enriquiment GO i KEGG

Mitjançant la funció *TCGAvisualize_EAbarplot* es va generar un fitxer que contenia diferents ontologies i vies canòniques significativament sobrerrepresentades pels gens expressats diferencialment. Aquestes es mostren en funció del seu *FDR ajustat* i de la proporció de gens trobats.

3.2.11. Gene Set Enrichment Analysis (GSEA)

Aquesta anàlisi d'enriquiment és un mètode computacional per a determinar si un conjunt de gens definit com a marcadors d'una via d'interès, mostra estadísticament diferències significatives i concordants entre les dues condicions establertes: pacients tractats amb quimioteràpia i pacients no tractats.

Es va generar una taula de valors que contenia informació sobre el nom del gen diferencialment expressat i el seu p-valor. Aquesta taula es va ordenar segons la significació dels p-valors (de més petits a més grans), i tot seguit, es va afegir una columna buida que posteriorment s'ompliria amb valors binaris. Es va buscar una signatura de senescència i una signatura de la via de 'Transforming growth factor beta (TGF-β)' (essent les signatures un conjunt de gens implicats en les vies). Les signatures es van entrar en forma de vector amb l'objectiu de mitjançant una estructura iterativa, omplir la columna buida amb 1s cada vegada que el gen del vector coincidís amb algun dels gens diferencialment expressats (vegeu l'esquematització del procés a la figura 6). Finalment, com que la taula es va ordenar segons la significació dels p-valors, es va realitzar un test hipergeomètric i es va interpretar amb un gràfic. En aquest gràfic s'hi representen els mil primers gens a l'eix de les 'x' i els seus p-valors a l'eix de les 'y', de manera que es va poder determinar si pels 1000 primers gens, hi havia més presència de gens expressats diferencialment implicats en la via d'interès, considerant-se així enriquits.

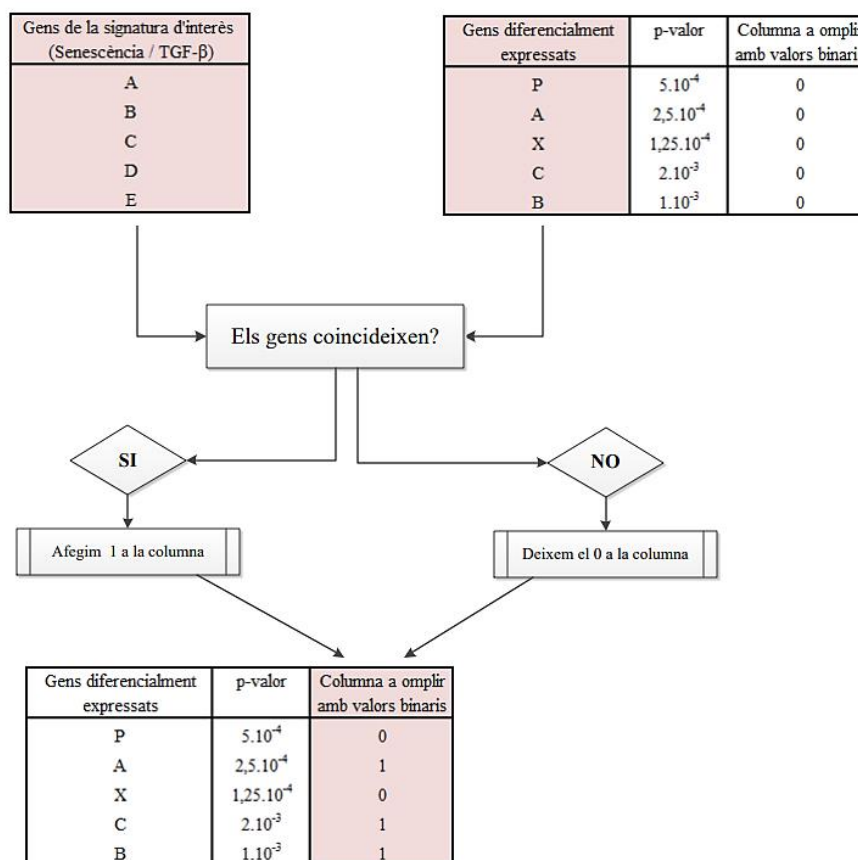


Figura 6. Esquematització del procés que es va seguir per a dur a terme l'anàlisi GSEA. En aquesta figura es representa un esquema del procés que es va seguir per a realitzar l'anàlisi d'enriquiment dels gens. Es marquen les columnes amb el color salmó per a indicar a quines d'elles es fa referència en cada moment (Font pròpia, 2020).

3.2.12. Anàlisi de supervivència

Es van realitzar anàlisis de supervivència comparant diferents variables. Primer, es van comparar els dos grups d'interès: pacients tractats amb quimioteràpia i pacients no tractats. Segon, es van comparar els estadis del tumor de tots els pacients, per així determinar quins són els estadis més agressius. I tercer, es van dividir els pacients tractats amb quimioteràpia en dos subgrups: pacients tractats amb alta expressió de biomarcadors de senescència i pacients tractats amb baixa expressió de biomarcadors de senescència. A més, es va realitzar la mateixa subdivisió per biomarcadors de la via TGF- β , ja que més endavant es menciona com aquest factor de creixement actua com a inductor en processos de senescència.

Amb l'objectiu d'estimar les probabilitats de supervivència dels instants en els quals ha succeït un esdeveniment, es va emprar el mètode de Kaplan-Meier, que utilitza la funció *survfit* de la llibreria *survival*. Aquesta funció s'aplica sobre un objecte *surv*, generant taules i gràfics de les corbes Kaplan-Meier. Per tal d'estimar si les corbes generades eren estadísticament significatives, es va emprar el *log-rank test* per mitjà de la funció *survdif*.

Finalment, es va realitzar el model de regressió de *Cox* amb la funció *coxp* amb la finalitat de determinar si les diferències entre els grups d'interès són significatives o es veuen esbiaixades per l'efecte d'altres variables com en serien l'estadi tumoral, el sexe, o l'edat dels pacients.

3.3. Estudi GSE65258

3.3.1. Dades d'expressió

En aquest cas, les dades es van obtenir del repositori GEO amb el nombre de referència GSE65258. Es tracta d'un perfil d'expressió gènica de 30 mostres de pulmó de ratolí. Aquestes mostres s'estructuren en tres condicions diferents: cèl·lules pulmonars normals (condició control), cèl·lules d'hiperplàsies pulmonars (lesions primerenques) i cèl·lules d'adenocarcinomes de pulmó (lesions avançades) (*Taula 2*).

Taula 2. Taula descriptiva de la informació continguda en el projecte GSE65256. La taula conté informació respecte al nombre total de mostres de l'estudi i el nombre de rèpliques de cada condició.

| Núm. mostres totals | Núm. de rèpliques de condició control | Núm. de rèpliques de lesions primerenques ('hiperplàstiques') | Núm. de rèpliques de lesions avançades ('full-blown adenocarcinoma') |
|---------------------|---------------------------------------|---|--|
| 30 | 16 | 7 | 7 |

Les anàlisis de dades de microarray es van realitzar mitjançant el software estadístic *R* i un conjunt de paquets disponibles de *Bioconductor*.

3.3.2. Adquisició i exploració de les dades a partir de la plataforma GEO

Prèviament a l'anàlisi, amb l'objectiu de realitzar un control de qualitat de les dades, aquestes es van descarregar crues des de la plataforma GEO gràcies a la funció *getGEOSuppFiles*. Posteriorment, es va procedir a la seva exploració utilitzant les funcions *class*, *dim* i *annotation*.

3.3.3. Control de qualitat

Es van examinar les imatges extretes del chip de microarray amb el propòsit de detectar anomalies que poguessin interferir en les intensitats de senyal. Per veure com es distribuïen les mostres per valors es va generar un histograma, mentre que per conèixer la distribució d'intensitats d'acord amb la mediana es va generar un boxplot.

3.3.4. Normalització

Amb el fi d'eliminar biaixos tècnics que interferissin en les intensitats de senyal, es van normalitzar les dades per quantils amb la funció *rma*. Posteriorment, es va procedir amb la sumarització dels valors de les diferents sondes per a obtenir-ne només un, i per últim es van transformar els valors a base logarítmica per a poder aplicar el test estadístic *t-test*.

3.3.5. Agregació de les dades

Es van crear clústers jeràrquics amb la funció *hclust* per a representar en dendrograms l'agrupació que seguien les mostres, detectant així possibles *outliers* en l'anàlisi. Es van definir

tant les distàncies basades en la correlació com el mètode de lincatge, en el nostre cas *war.D2*, *complete* i *average*. Es va repetir aquest procés però mitjançant la tècnica de la reducció de la dimensió amb l'ús de la funció *pca*. Aquesta, representa totes les variables inicials en tres i això permet poder-les representar i combinar els gens per a explorar al màxim la variabilitat.

3.3.6. Anàlisi d'expressió diferencial

Es va aplicar el test estadístic *t-test* per tal de comparar les dues condicions: lesions primerenques i lesions avançades. Aquesta anàlisi afegeix un terme de moderació al denominador per a corregir la limitació en el nombre de mostres. En aquest estudi es defineix el disseny de la matriu emprant la funció *makeContrasts* i *eBayes* del paquet de *limma*. A continuació, es va realitzar el procés d' anotació amb la funció *mget* i la base de dades *mouse4302.db*, convertint així l'identificador del gen en el seu nom complet. Amb l'objectiu de representar els nivells d'expressió dels gens, es va generar un *HeatMap* per mitjà de la funció *heatmap.2*, i un *volcanoplot* amb la funció *with* i *subset*.

3.3.7. Gene Set Enrichment Analysis (GSEA)

Aquesta anàlisi d'enriquiment és un mètode computacional per a determinar si un conjunt de gens definit com a marcadors d'una via d'interès, mostra estadísticament diferències significatives i concordants entre les dues condicions establertes: pacients tractats amb quimioteràpia i pacients no tractats.

Es va crear una taula de valors que contenia informació sobre el nom del gen diferencialment expressat i el seu p-valor. Aquesta taula es va ordenar segons la significació dels p-valors (de més petits a més grans), i tot seguit, es va afegir una columna buida que posteriorment s'ompliria amb valors binaris. Es va buscar una signatura de senescència (essent la signatura un conjunt de gens implicats en la via), i es van entrar en forma de vector amb l'objectiu de mitjançant una estructura iterativa, omplir la columna buida amb 1s cada vegada que el gen del vector coincidís amb algun dels gens diferencialment expressats (vegeu l'esquematització del procés a la *figura 6*). Finalment, com que la taula es va ordenar segons la significació dels p-valors, es va realitzar un test hipergeomètric i es va interpretar amb un gràfic. En aquest gràfic s'hi representen els mil primers gens a l'eix de les 'x' i els seus p-valors a l'eix de les 'y', de manera que es va poder determinar si pels 1000 primers gens, hi havia més presència de gens expressats diferencialment implicats en la via d'interès, considerant-se així enriquits.

4. Resultats i discussió

4.1. Estudi TCGA-LUAD

En el present estudi, es volen analitzar les pautes d'expressió gènica dels tumors dels pacients que han rebut quimioteràpia, i concloure si presenten una signatura de senescència significativament superior comparat amb els tumors dels pacients que no han rebut quimioteràpia. A més, es vol analitzar l'efecte de la quimioteràpia sobre la supervivència dels pacients amb càncer de pulmó, discernint entre els que han rebut un tractament farmacològic i els que no han rebut cap mena de tractament. Finalment, es vol trobar una relació entre la possible signatura de senescència en tumors i la supervivència dels pacients.

4.1.1. Anàlisi d'expressió diferencial

Es van determinar els gens que s'expressaven diferencialment classificant les dades segons els dos grups d'estudi: els tractats mitjançant quimioteràpia i els no tractats. Es van adquirir un total de 522 casos, inclosos 242 pacients masculins i 280 femenins, dels quals 109 s'havien sotmès a tractament farmacològic i 367 no havien rebut tractament. Es van analitzar un total de 14.889 gens. La *taula 3* mostra els trenta gens més diferencialment expressats entre els tractats i no tractats, mentre que la *imatge 18* de l'annex conté una llista més exhaustiva de gens diferencialment expressats entre les dues condicions. Concretament, es van filtrar considerant un p-valor menor a 0.05 i un logaritme del 'Fold-Change' (log fold-change, logFC) en valor absolut major que 1. Obtenint així, un total de 198 gens diferencialment expressats entre les dues condicions (*imatge 18* de l'annex).

Tal com s'observa tant a la *taula 3* com a la *imatge 18* de l'annex, els valors de logFC poden ser negatius o positius, indicant la direcció de l'expressió, és a dir, si el gen es troba infraexpressat o sobreexpressat. Nogensmenys, si s'observa amb més detall la taula completa dels gens diferencialment expressats de l'annex, s'hi poden detectar més valors positius que negatius.

Taula 3. Taula del resultat de l'anàlisi d'expressió diferencial del projecte TCGA-LUAD. Les files representen els gens i les columnes representen els seus valors pertinents de 'log fold-change (logFC)', log counts-per-million (logCPM), log fold-change ratio (LR), p-valor (p-value) i p-valor ajustat (FDR). Els valors negatius corresponen a una infraexpressió del gen en pacients tractats amb quimioteràpia, mentre que els valors positius en representen una sobreexpressió. En aquesta taula només apareixen els trenta primers gens per tal de poder visualitzar els seus valors, si es desitjés veure la taula del resultat complet dirigiu-vos a la imatge 18 de l'annex.

| | logFC | logCPM | LR | PValue | FDR |
|-----------|-------------------|-------------------|------------------|----------------------|----------------------|
| MSMP | -2.24759776539053 | 2.53427652842193 | 156.417288443507 | 6.86248783961373e-36 | 1.02244206325385e-31 |
| CUL4A | -1.06465604610093 | 6.81243144388149 | 151.412693764143 | 8.51506218229911e-35 | 6.34329557270373e-31 |
| PI3 | -2.96641444564115 | 4.10020084552877 | 138.067239722223 | 7.04504968394237e-32 | 3.49880650803525e-28 |
| CYP2B6 | -2.94108607878637 | 0.731454556957044 | 130.475100507965 | 3.22552996608553e-30 | 1.20142927411771e-26 |
| NPY | -6.79953637924709 | 2.43609353890901 | 109.937792979765 | 1.01114586439642e-25 | 2.51084370560705e-22 |
| CALB1 | -4.56291617020398 | 4.29265334691094 | 97.2191269629668 | 6.20642170625858e-23 | 1.32099252859352e-19 |
| MSTN | 4.14540510022863 | 1.45010436073434 | 93.8595229411857 | 3.38730727289179e-22 | 6.30843638235185e-19 |
| DPCR1 | 4.08611643239507 | 7.00925364334106 | 92.9515265639079 | 5.3590950084859e-22 | 8.87168405904793e-19 |
| LEFTY1 | -1.68517317103952 | 0.690096420204387 | 88.640891967978 | 4.73401304243856e-21 | 7.05320603192921e-18 |
| CACNA1A | -1.72422891922404 | 0.609805001215126 | 86.1500898686054 | 1.66784341763407e-20 | 2.25901809812091e-17 |
| CTCF | -1.4309279832316 | 0.396439014796355 | 80.9050738517274 | 2.36824826195233e-19 | 2.94037757123565e-16 |
| PAX8 | -1.35190058506319 | 2.5248061479498 | 75.4173312708195 | 3.81030715657002e-18 | 4.36690510197975e-15 |
| ALB | 5.98281111324178 | 5.97261076430484 | 69.9239032226404 | 6.16369902847712e-17 | 6.15412629561553e-14 |
| LOC283867 | 2.65367585821253 | 0.990977343818819 | 69.9136424206467 | 6.19584498518243e-17 | 6.15412629561553e-14 |
| DUOX2 | 2.19447838023063 | 4.34544479865157 | 60.8192811759817 | 6.25621597961603e-15 | 5.8257101175187e-12 |
| KRT13 | -2.28889843932713 | 3.41031878159801 | 59.0605218782498 | 1.5289619041647e-14 | 1.34000020059705e-11 |
| NLGN4X | -1.1642057600513 | 1.7439571918562 | 55.8816592545992 | 7.69672523838354e-14 | 5.73367546633382e-11 |
| LRRC55 | 1.9092269078089 | 1.59478451851847 | 52.4369523748004 | 4.4428498058535e-13 | 2.87800083727875e-10 |
| TCHH | -1.28866280287277 | 0.952920073963788 | 51.2846985854953 | 7.98959687442259e-13 | 4.95987515966759e-10 |
| HIST1H4J | 1.90768278634945 | 2.86665483780127 | 51.1752681604572 | 8.44762574158446e-13 | 4.96214923993344e-10 |
| CUZD1 | -1.19868937002656 | 1.28223264737708 | 51.1266722912167 | 8.6593650740499e-13 | 4.96214923993344e-10 |
| ALDOB | 2.28966911626815 | 1.64478474476011 | 46.413107479934 | 9.57726979252618e-12 | 4.92040491858095e-09 |
| DUOX2 | 1.76848492828998 | 2.02648613692112 | 45.9074231597058 | 1.2397529284201e-11 | 6.15702629351034e-09 |
| KRT86 | -1.15732727408924 | 1.92166085046035 | 45.2843394007788 | 1.70406313809745e-11 | 8.18994732081095e-09 |
| ZBTB20 | 1.54026595268717 | 1.09577129393768 | 44.80934457156 | 2.17183738596029e-11 | 1.01058907794196e-08 |
| FABP7 | 4.34252464719026 | 2.55324596391505 | 43.2647592616652 | 4.7811758135463e-11 | 2.0951393660596e-08 |
| CGB5 | -3.18586481544171 | 1.14929170932152 | 42.9246807981098 | 5.68884126139685e-11 | 2.42165845581570e-08 |
| HIST4H4 | 1.52313129327655 | 1.39793119711732 | 42.6539470135772 | 6.53324661038386e-11 | 2.70385670133636e-08 |
| COL11A2 | 1.91683879990461 | 2.60328398449254 | 42.3812936693494 | 7.51049530250284e-11 | 3.02429377059432e-08 |
| HIST2H3C | 2.5135785122493 | 2.29315136080686 | 39.9731433148637 | 2.57478747578386e-10 | 9.83634835941122e-08 |

Es van representar els nivells d'expressió del total de gens (14.889 gens) per mitjà d'un 'volcano plot' (figura 7). En aquest gràfic s'hi pot observar molta més quantitat de gens amb p-valors estadísticament significatius sobreexpressats en els pacients tractats amb quimioteràpia, rondant valors de logFC d'entre 1 i 3; mentre que els que estan infraexpressats en aquesta condició, romanen en un rang d'entre 1 i 2.

D'altra banda es va fer una representació amb un 'Heat Map' (figura 8) dels primers 17 gens diferencialment més expressats amb l'objectiu de poder veure les diferències d'expressió amb més claredat. Així doncs, aquest gràfic tot i presentar bastants punts d'infraexpressió (verds), s'hi observa una tendència majoritària en els valors corresponents a sobreexpressió (vermells), el que ens confirma els resultats obtinguts en el volcano plot.

En definitiva, aquestes evidències indiquen que la majoria dels gens es troben sobreexpressats en els pacients que es van sotmetre a quimioteràpia comparat amb els que no van rebre cap mena de tractament.

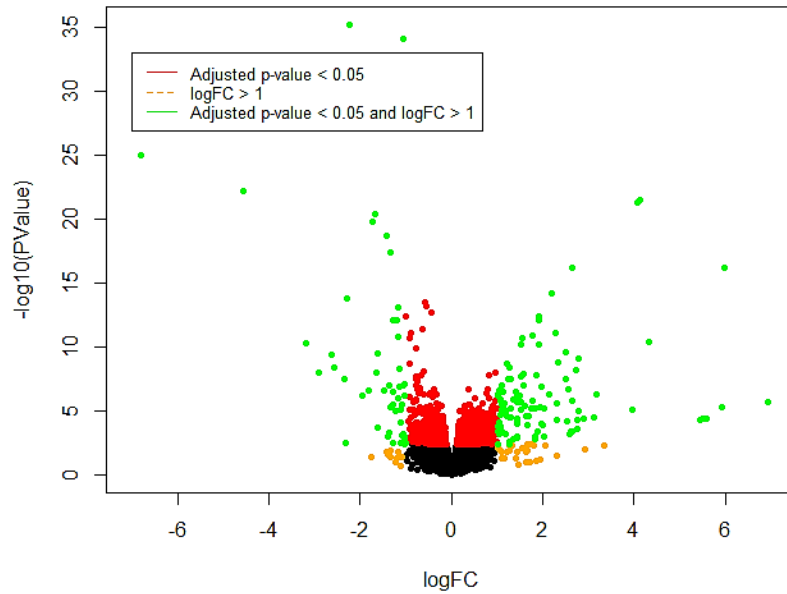


Figura 7. Volcano plot representant els gens diferencialment expressats del projecte TCGA-LUAD. Els punts vermells signifiquen aquells valors els quals el seu p-value és menor a 0.05. Els taronges són aquells els quals el seu log Fold-Change (logFC) és major a 1 i finalment, els verds són aquells als quals els corresponen les dues condicions.

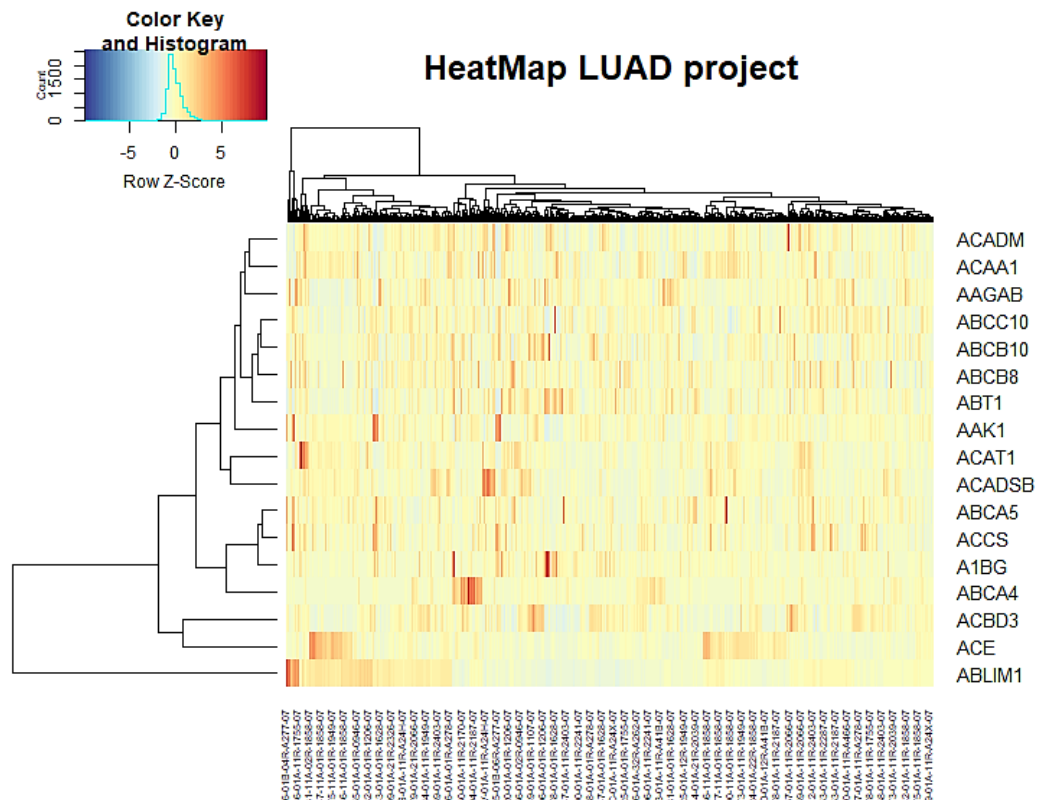


Figura 8. HeatMap representant els gens diferencialment expressats del projecte TCGA-LUAD. Les files representen els gens i les columnes representen les mostres. Els nivells d'expressió en aquest gràfic, tal com s'indica a la llegenda estan codificats per la intensitat de color, essent el color vermell l'indicador d'alta expressió i el blau l'indicador de baixa expressió. Noteu que només s'ha anotat el nom dels primers 17 gens per tal que en el gràfic les diferències d'expressió fossin visibles, i que el nom de les mostres no es veu complet ja que aquests eren massa llargs.

4.1.2. Anàlisi d'enriquiment GO i KEGG

Amb el propòsit d'extreure conclusions sobre el conjunt de gens expressats diferencialment, es va realitzar una anàlisi d'enriquiment per 'Gene Ontology (GO)' i 'Kyoto Encyclopedia of Genes and Genomes (KEGG)'. En aquestes, es volien determinar en quins processos biològics, components cel·lulars, funcions moleculars i vies es trobaven sobrerrepresentats aquest conjunt de gens.

D'una banda, es van considerar sobrerrepresentats els processos de percepció sensorial de l'olfacte, probablement degut al tabaquisme, la percepció sensorial d'estímuls químics, causat possiblement per la quimioteràpia aplicada basada en cisplatí, i en processos del cicle cel·lular i mitòtic entre d'altres (*figura 9A*). Aquesta última evidència és de molt interès, ja que molts dels biomarcadors de senescència (com p16, p21 i p53) pertanyen a proteïnes que intervien en aquests dos cicles. Això ens podria indicar que la majoria de gens que s'han trobat diferencialment expressats, podrien presentar una sobrerrepresentació en processos biològics que intervien en les vies de senescència.

D'altra banda, es van estimar sobrerrepresentats la majoria de components cel·lulars del lumen i del mitocondri entre els més rellevants (*figura 9B*). El lumen es troba normalment a les membranes d'orgànuls, i és responsable de la producció d'hormones i productes secretoris. A més, el mitocondri és un dels orgànuls principals encarregats d'intervenir en processos d'envelliment com la senescència.

Tal com s'esmenta a l'apartat 1.2.2. *SASP*, les cèl·lules senescentes es caracteritzen per secretar de manera paracrina aquest complex proinflamatori anomenat *SASP*, de manera que aquesta evidència, juntament amb la sobrerrepresentació de components del mitocondri, ens podria indicar que la major part dels gens diferencialment expressats tenen sobrerrepresentats components cel·lulars que podrien interferir en tant la secreció del *SASP* com en la inducció de processos de senescència.

Per la funció molecular, es van detectar sobrerrepresentades les funcions associades a l'activitat receptora de l'olfacte, probablement degut al tabaquisme com hem esmentat anteriorment, i les funcions acoblades a la unió i transició d'ions i metalls entre d'altres (*figura 9C*). Dins del camp de la senescència i la seva investigació, s'ha reportat una relació entre la inflamació i la secreció de *SASP*, ja que aquest és un conjunt de citocines, quimosines i factors de creixement inflamatoris. Com anteriorment s'ha comentat als apartats 1.2.2. *SASP* i 1.2.3. *SASP i càncer*, la secreció de *SASP* augmenta la producció intracel·lular d'espècies d'oxigen radioactiu (ROS) estimulants l'afluència de Ca^{2+} , fet que indirectament induïx l'alliberació d'interleucines proinflamatòries.

Aquesta evidència ens podria estar indicant que una bona part dels gens reportats com a diferencialment expressats en pacients sota tractament farmacològic, tenen una funció molecular sobrerrepresentada vinculada a processos de senescència i secreció del *SASP*.

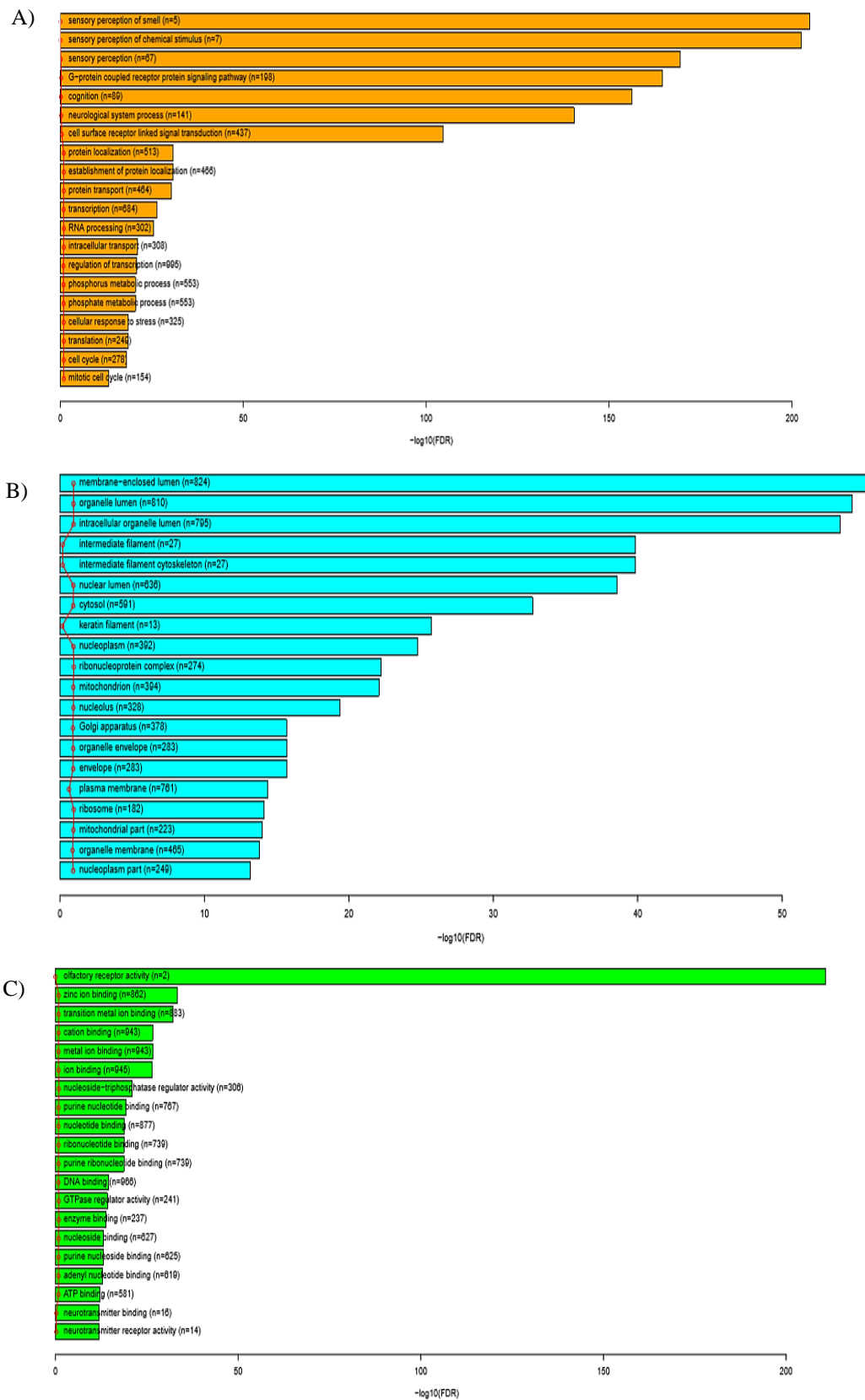


Figura 9. Resultats de l'anàlisi d'enriquiment 'Gene Ontology (GO)' del conjunt de gens expressats diferencialment del projecte TCGA-LUAD. A) Anàlisi d'enriquiment segons el procés biològic. B) Anàlisi d'enriquiment segons el component cel·lular. C) Anàlisi d'enriquiment segons la funció molecular. Per tal de fer més amena la visualització dels gràfics només es mostren els vint primers.

Finalment, les vies que es van trobar sobrerrepresentades van ser les relacionades amb els mecanismes moleculars del càncer -cosa que esperàvem, ja que les mostres són tumorals- i les vies de senyalització de mTOR i disfunció mitocondrial entre d'altres (*figura 10*). La via de senyalització de mTOR és una de les més afectades quan s'ha desenvolupat un procés cancerigen. La disfunció mitocondrial podria indicar una relació amb els resultats que hem obtingut anteriorment, on s'esmenta que la inflamació i secreció del SASP desestabilitzen l'aflluència de Ca^{2+} , i aquest, que es conté al mitocondri podria estar induint una major producció de ROS, i per conseqüència, de senescència.

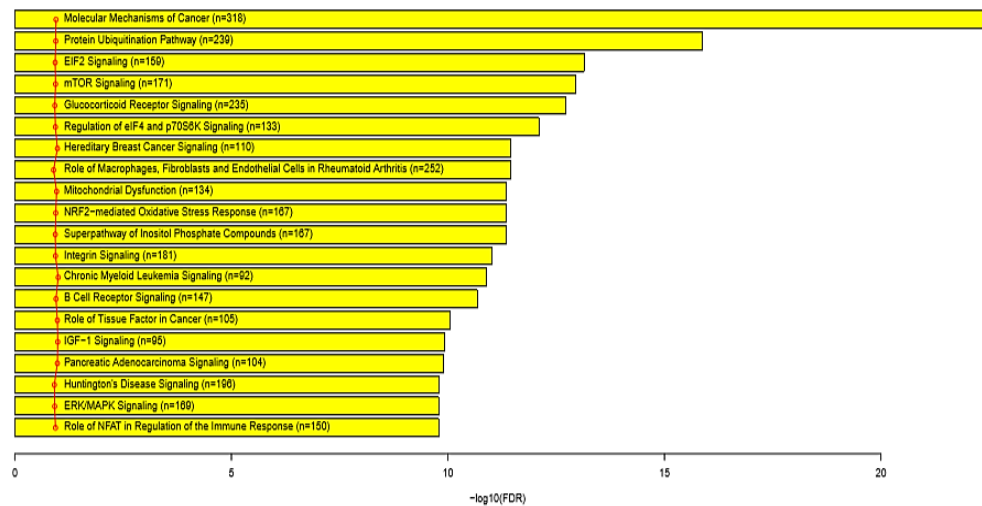


Figura 10. Resultats de l'anàlisi d'enriquiment 'Kyoto Encyclopedia of Genes and Genomes (KEGG)' del conjunt de gens expressats diferencialment del projecte TCGA-LUAD. L'anàlisi d'enriquiment es va fer segons la via a la que pertanyen els gens. Per tal de fer més amena la visualització del gràfic només es mostren els vint primers gens.

En definitiva, podem concloure que la major part dels gens diferencialment expressats en pacients tractats amb quimioteràpia, pateixen una sobrerrepresentació en processos biològics, components cel·lulars, funcions moleculars i vies que estan vinculades a processos de senescència.

4.1.3. Gene Set Enrichment Analysis (GSEA)

Per tal de determinar si els processos de senescència i de TGF- β es troben enriquits en la progressió i el pronòstic del càncer de pulmó, es van analitzar el conjunt de gens considerats com a diferencialment expressats i es van comparar amb les signatures pertinents: La signatura de senescència procedia de la pàgina oficial de GSEA (<https://www.gsea-msigdb.org/gsea>) realitzada per Fridman, AL i Tainsky, MA i conté 77 gens sobreexpressats en processos de senescència, i la de TGF- β es va extreure de la pàgina oficial de PANTHER (<http://www.pantherdb.org/>) i conté 88 gens sobreexpressats en la via de senyalització d'aquest factor de creixement.

Es va realitzar un test d'enriquiment basat en la distribució hipergeomètrica sobre la seqüència de gens (ordenats segons el seu nivell d'expressió diferencial) entre pacients tractats i no tractats. Per una banda, es va representar el resultat de l'anàlisi per la signatura de senescència, amb un gràfic convencional mostrant els mil primers gens a l'eix de les 'x', ordenats en successió, i el seu p-valor del test d'enriquiment corresponent, a l'eix de les 'y' (*figura 11*).

L'indicador que defineix si els gens estan enriquits és el p-valor, ja que com més petit és aquest més significatiu són els resultats. En altres paraules, perquè els gens es trobessin enriquits, els p-valors més petits i per tant més significatius, haurien d'estar al capdavant del gràfic. Tal com s'observa a la figura, els gens diferencialment expressats no es troben enriquits en processos de senescència perquè pels primers 600 gens, els p-valors respectius estan al voltant d'1.

D'altra banda, es va representar de la mateixa manera el resultat de l'anàlisi per la signatura de TFG- β (figura 12). En aquest, els gens diferencialment expressats sí que es troben enriquits en la via, ja que com s'observa a la figura, els p-valors més petits i per tant, més significatius, es troben liderant el gràfic, la qual cosa denota un enriquiment dels gens per a la via d'aquest factor de creixement.

S'han reportat alguns dels efectes citostàtics de TGF- β , i entre ells s'inclou la inducció de p15 i p21 que s'associa amb la promoció de la senescència. Aquest factor de creixement es vincula a múltiples accions reguladores, com la producció d'espècies reactives de l'oxigen (ROS), i aquest també és un promotor de processos de senescència. A més, el SASP produeix la secreció de diverses molècules de senyalització, i entre elles s'inclou aquest factor de creixement, de forma que podria estar induint i mantenint el fenotip senescent.

Al contrari de les nostres expectatives, aquests resultats no mostren cap evidència que hi hagi una signatura de senescència als tumors de pacients tractats amb quimioteràpia, i per tant, no podem concloure cap relació entre la via de senyalització de TGF- β i la senescència en els tumors.

Com a reflexió especulativa i personal, els resultats podrien indicar que el fet d'estar tractant dades d'adenocarcinomes (estadis primerencs del tumor) i no carcinomes, el TFG- β mitjançant les accions reguladores esmentades anteriorment, podria estar promovent l'acumulació de cèl·lules senescentes perquè aquestes secretin el complex SASP, i per conseqüència, se secreti més TFG- β amb l'objectiu de mantenir aquest fenotip senescent, i poder afectar les cèl·lules veïnes perquè també esdevinguin senescentes.

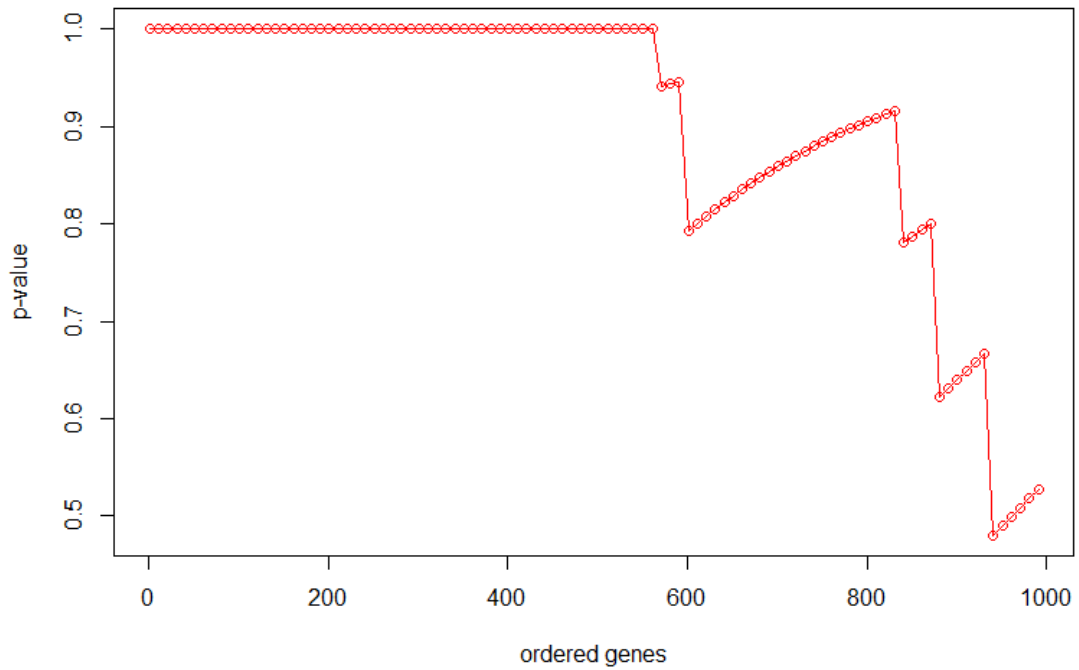


Figura 11. Resultats de l'anàlisi d'enriquiment de GSEA per una signatura de senescència en pacients amb càncer de pulmó del projecte TCGA-LUAD. Com es pot observar al gràfic, els gens es troben representats en una successió de dos-cents en dos-cents. El seu p-valor respectiu es troba representat a l'eix de les "y", de manera que com més petit sigui aquest, més significatiu és el resultat.

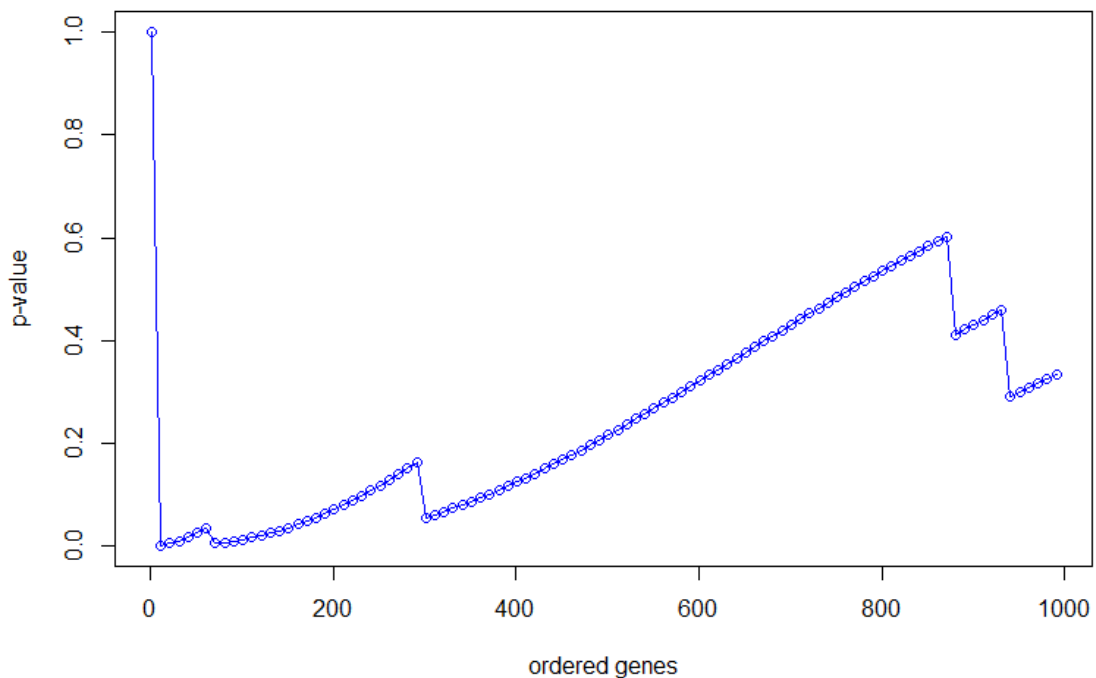
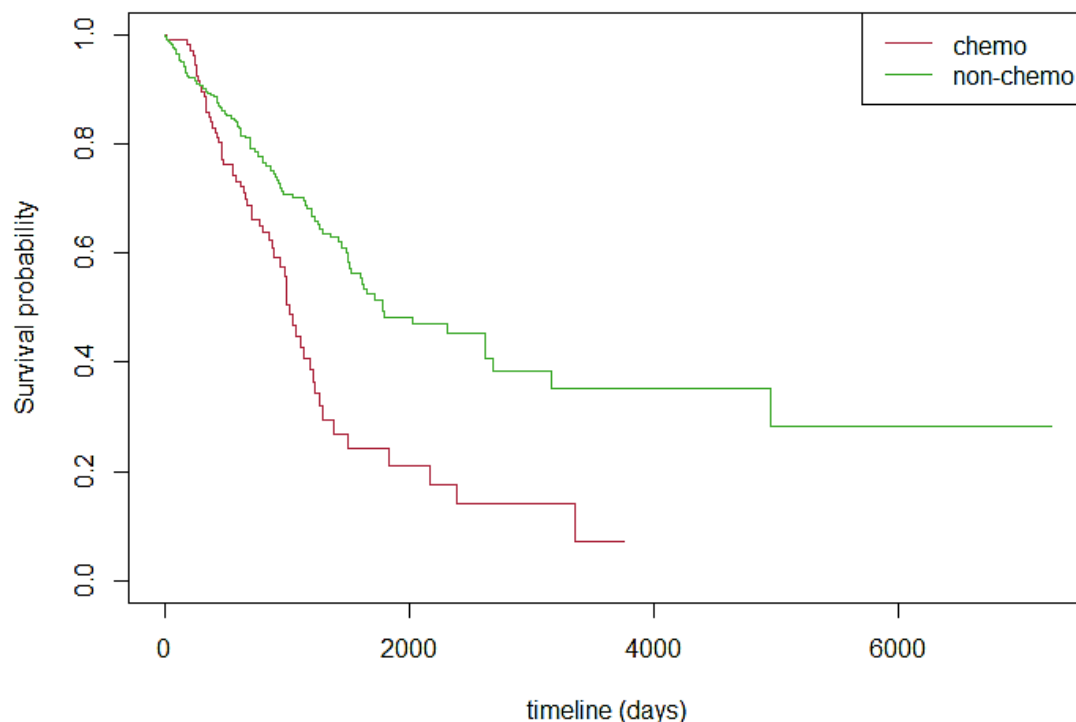


Figura 12. Resultats de l'anàlisi d'enriquiment de GSEA per una signatura de TGF- β en pacients amb càncer de pulmó del projecte TCGA-LUAD. Com es pot observar al gràfic, els gens es troben representats en una successió de dos-cents en dos-cents. El seu p-valor respectiu es troba representat a l'eix de les "y", de manera que com més petit sigui aquest, més significatiu és el resultat.

4.1.4. Anàlisi de supervivència

Les anàlisis de supervivència permeten investigar el temps fins que un esdeveniment succeeixi. Per a aquestes es va utilitzar el mètode de Kaplan-Meier, per així poder estudiar la supervivència de les variables d'interès. Aquest, és un mètode no paramètric que s'utilitza per a estimar la probabilitat de supervivència a partir dels temps de supervivència observats. Així doncs, el gràfic que es genera amb aquest mètode, ens indica la probabilitat de supervivència contra el temps (tal com s'observa a les figures que s'exposen posteriorment). També es va utilitzar la regressió de Cox, la qual permet estudiar conjuntament l'efecte de diversos factors o variables sobre la supervivència.

A la figura 13, es pot observar com la supervivència es veu afectada en aquells pacients que estaven sota tractament farmacològic, confirmant així, que la probabilitat de supervivència dels pacients que no es van sotmetre a cap mena de tractament, és molt menor comparada amb la dels pacients que es van sotmetre a quimioteràpia.



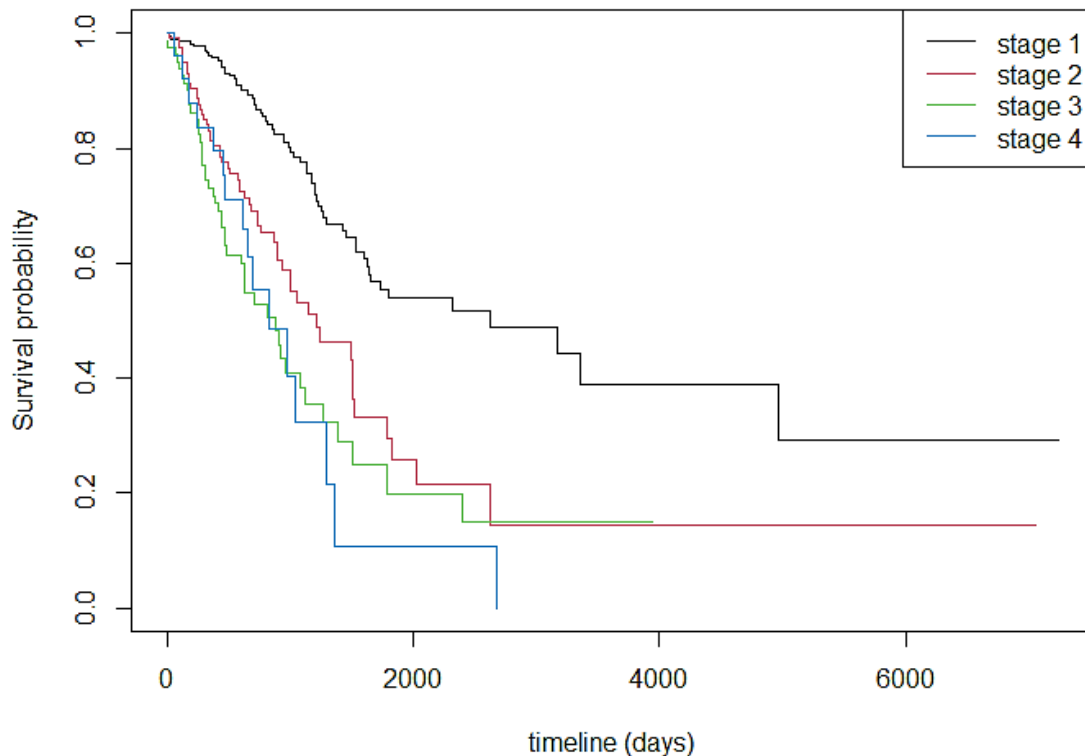
n=467, 55 observations deleted due to missingness.

| | N | Observed | Expected | (O-E) ² /E | (O-E) ² /V |
|---------------|-----|----------|----------|-----------------------|-----------------------|
| data\$treat=1 | 109 | 60 | 36.8 | 14.5 | 18.8 |
| data\$treat=2 | 358 | 111 | 134.2 | 4.0 | 18.8 |

Chisq= 18.8 on 1 degrees of freedom, p= 1e-05

Figura 13. Corba de supervivència Kaplan-Meier de pacients amb càncer de pulmó del projecte TCGA-LUAD segons el tractament aplicat. Es representa la probabilitat de supervivència a l'eix de les "y" en funció del temps a l'eix de les "x". La corba vermella correspon als pacients tractats amb quimioteràpia, mentre que la verda correspon als pacients que no han estat tractats. S'afegeix el test estadístic per a comprovar si les diferències entre les corbes són significatives, en aquest cas si ho són perquè el p-valor és més petit que 0.05. Dins del projecte hi havia valors els quals no es va reportar quin era el tractament que es va subministrar al pacient. Aquests han sigut eliminats i no han sigut considerats per l'anàlisi.

Amb l'objectiu de corroborar que els tumors d'estadis avançats estan associats a una pitjor prognosi, igual que a una taxa de supervivència menor, es va realitzar una anàlisi de supervivència entre els estadis en què es trobaven els tumors (*figura 14*). Com s'observa en el gràfic, els que estan en l'estadi quatre tenen menys probabilitat de supervivència que els de l'estadi u.



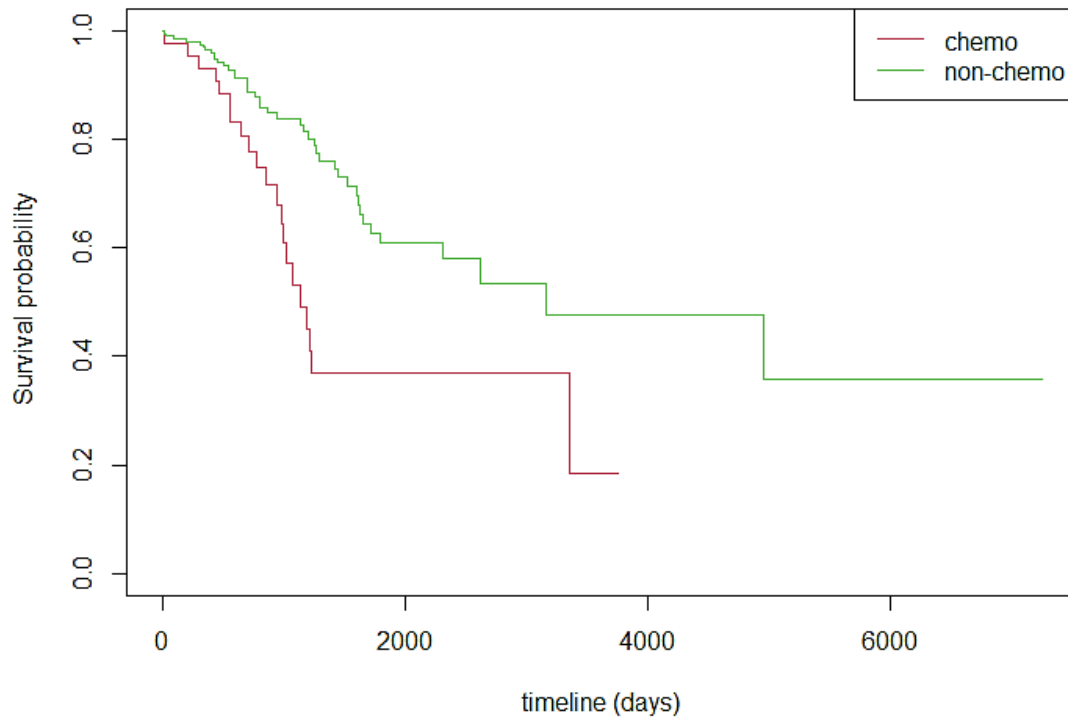
n=505, 17 observations deleted due to missingness.

| | N | Observed | Expected | (O-E) ² /E | (O-E) ² /V |
|-------------------|-----|----------|----------|-----------------------|-----------------------|
| clin_df1\$stage=1 | 275 | 66 | 112.92 | 19.49 | 52.29 |
| clin_df1\$stage=2 | 122 | 54 | 39.04 | 5.74 | 7.36 |
| clin_df1\$stage=3 | 82 | 46 | 22.69 | 23.96 | 27.53 |
| clin_df1\$stage=4 | 26 | 16 | 7.36 | 10.14 | 10.61 |

chisq= 60.4 on 3 degrees of freedom, p= 5e-13

Figura 14. Corba de supervivència Kaplan-Meier de pacients amb càncer de pulmó del projecte TCGA-LUAD segons l'estadi del tumor reportat. Es representa la probabilitat de supervivència a l'eix de les "y" en funció del temps a l'eix de les "x". La corba negra correspon a l'estadi 1 del tumor, la corba vermella correspon a l'estadi 2 del tumor, la corba verda correspon a l'estadi 3 del tumor i la corba de color blau correspon a l'estadi 4 del tumor. S'afegeix el test estadístic per a comprovar si les diferències entre les corbes són significatives, en aquest cas si que ho són perquè el p-valor és més petit que 0.05. Dins del projecte hi havia valors els quals no es va reportar quin era l'estadi del tumor del pacient. Aquests han sigut eliminats i no han sigut considerats per l'anàlisi.

També hem analitzat l'efecte del tractament estratificat segons l'estadi del tumor (*figures 15, 16, 17, 18*). Podem observar que l'efecte del tractament sobre la supervivència està condicionat per l'estadi del tumor. En l'estadi 1 veiem clarament una supervivència menor entre els pacients tractats amb quimioteràpia mentre que en l'estadi 4 el comportament és al contrari i els pacients tractats amb quimioteràpia tenen major supervivència (encara que les diferències no siguin significatives).

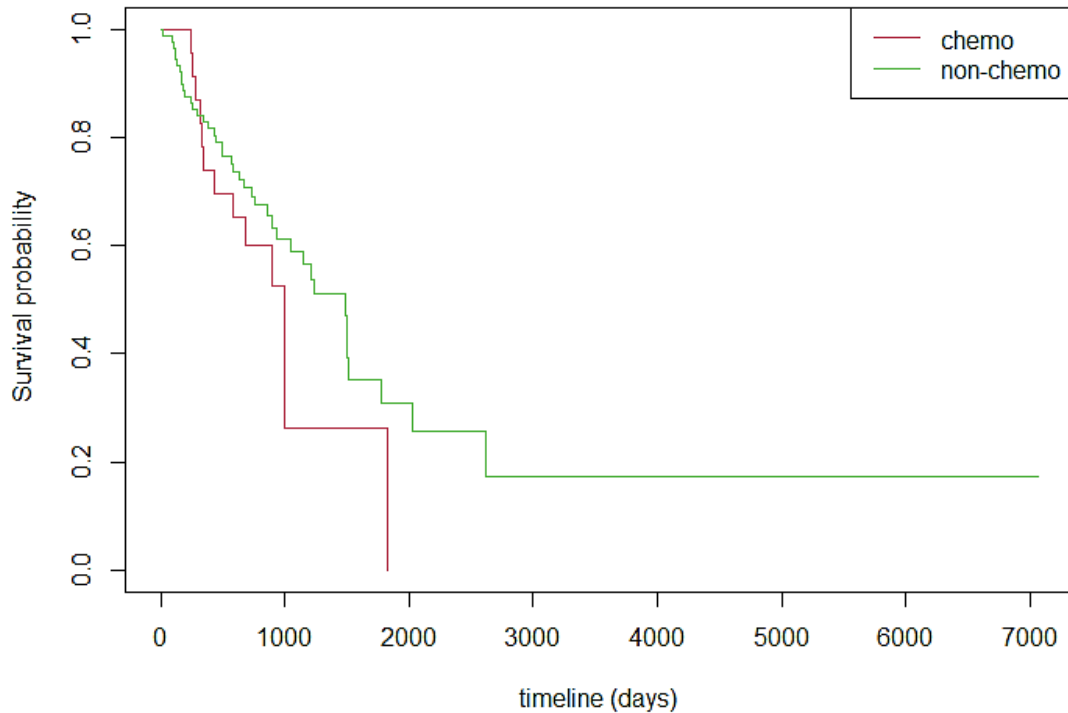


n=253, 34 observations deleted due to missingness.

| | N | Observed | Expected | (O-E) ² /E | (O-E) ² /V |
|---------------|-----|----------|----------|-----------------------|-----------------------|
| data\$treat=1 | 44 | 21 | 11 | 9.14 | 11.2 |
| data\$treat=2 | 209 | 42 | 52 | 1.93 | 11.2 |

ChiSq= 11.2 on 1 degrees of freedom, p= 8e-04

Figura 15. Corba de supervivència Kaplan-Meier de pacients amb càncer de pulmó del projecte TCGA-LUAD estratificada per l'estadi 1. Es representa la probabilitat de supervivència a l'eix de les "y" en funció del temps a l'eix de les "x". La corba negra correspon a l'estadi 1 del tumor, la corba vermella correspon a l'estadi 2 del tumor, la corba verda correspon a l'estadi 3 del tumor i la corba de color blau correspon a l'estadi 4 del tumor. S'afegeix el test estadístic per a comprovar si les diferències entre les corbes són significatives, en aquest cas si que ho són perquè el p-valor és més petit que 0.05. Dins del projecte hi havia valors els quals no es va reportar quin era l'estadi del tumor del pacient. Aquests han sigut eliminats i no han sigut considerats per l'anàlisi.

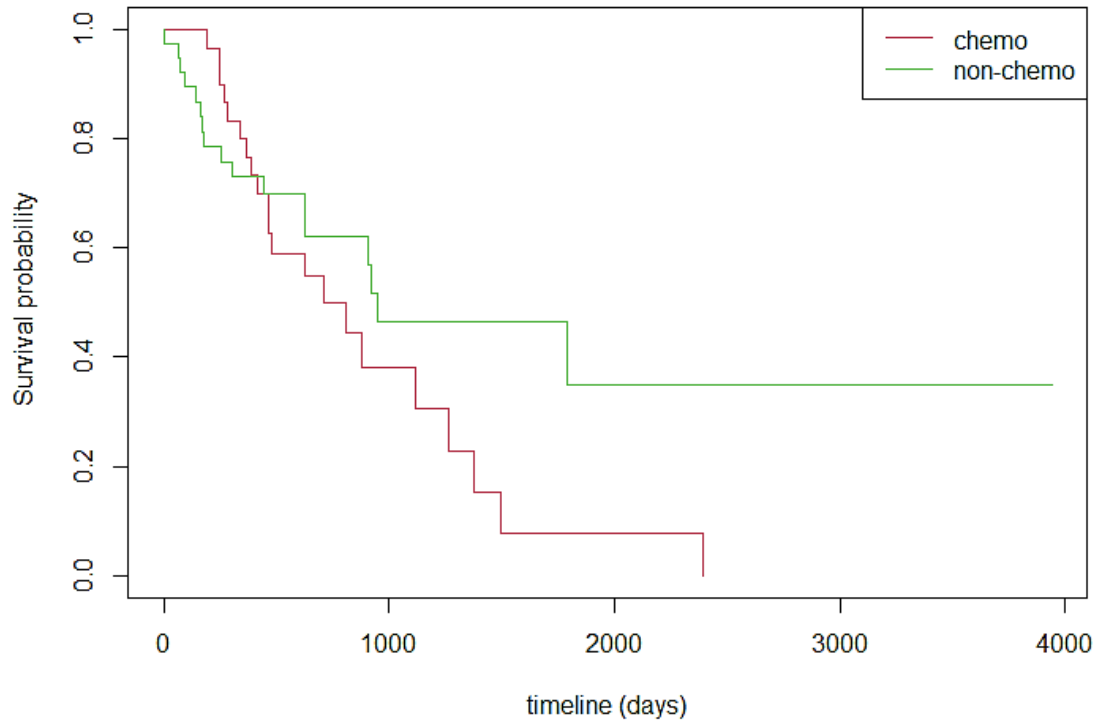


n=115, 17 observations deleted due to missingness.

| | N | Observed | Expected | (O-E) ² /E | (O-E) ² /V |
|---------------|----|----------|----------|-----------------------|-----------------------|
| data\$treat=1 | 24 | 13 | 9.29 | 1.480 | 1.85 |
| data\$treat=2 | 91 | 40 | 43.71 | 0.315 | 1.85 |

Chisq= 1.9 on 1 degrees of freedom, p= 0.2

Figura 16. Corba de supervivència Kaplan-Meier de pacients amb càncer de pulmó del projecte TCGA-LUAD estratificada per l'estadi 2. Es representa la probabilitat de supervivència a l'eix de les "y" en funció del temps a l'eix de les "x". La corba negra correspon a l'estadi 1 del tumor, la corba vermella correspon a l'estadi 2 del tumor, la corba verda correspon a l'estadi 3 del tumor i la corba de color blau correspon a l'estadi 4 del tumor. S'afegeix el test estadístic per a comprovar si les diferències entre les corbes són significatives, en aquest cas no ho són perquè el p-valor és major que 0.05. Dins del projecte hi havia valors els quals no es va reportar quin era l'estadi del tumor del pacient. Aquests han sigut eliminats i no han sigut considerats per l'anàlisi.

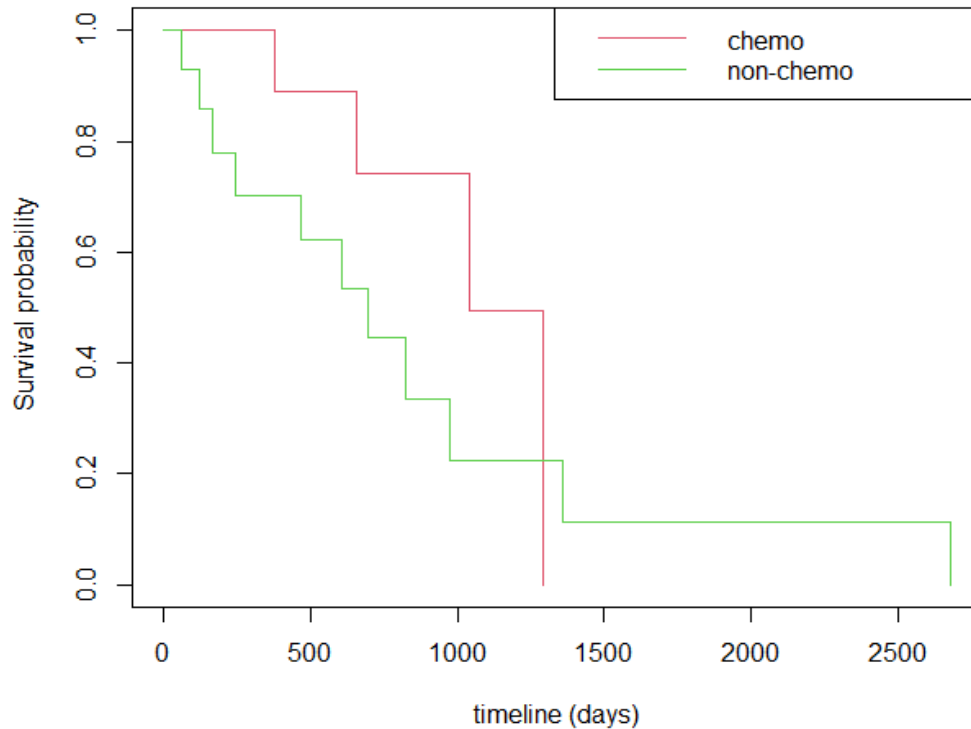


n=69, 24 observations deleted due to missingness.

| | N | Observed | Expected | $(O-E)^2/E$ | $(O-E)^2/V$ |
|---------------|----|----------|----------|-------------|-------------|
| data\$treat=1 | 31 | 21 | 16.5 | 1.235 | 2.24 |
| data\$treat=2 | 38 | 17 | 21.5 | 0.947 | 2.24 |

chisq= 2.2 on 1 degrees of freedom, p= 0.1

Figura 17. Corba de supervivència Kaplan-Meier de pacients amb càncer de pulmó del projecte TCGA-LUAD estratificada per l'estadi 3. Es representa la probabilitat de supervivència a l'eix de les "y" en funció del temps a l'eix de les "x". La corba negra correspon a l'estadi 1 del tumor, la corba vermella correspon a l'estadi 2 del tumor, la corba verda correspon a l'estadi 3 del tumor i la corba de color blau correspon a l'estadi 4 del tumor. S'afegeix el test estadístic per a comprovar si les diferències entre les corbes són significatives, en aquest cas no ho són perquè el p-valor és major que 0.05. Dins del projecte hi havia valors els quals no es va reportar quin era l'estadi del tumor del pacient. Aquests han sigut eliminats i no han sigut considerats per l'anàlisi.



n=23, 11 observations deleted due to missingness.

| | N | Observed | Expected | (O-E) ² /E | (O-E) ² /V |
|---------------|----|----------|----------|-----------------------|-----------------------|
| data\$treat=1 | 9 | 4 | 6.04 | 0.691 | 1.31 |
| data\$treat=2 | 14 | 11 | 8.96 | 0.467 | 1.31 |

Chisq= 1.3 on 1 degrees of freedom, p= 0.3

Figura 18. Corba de supervivència Kaplan-Meier de pacients amb càncer de pulmó del projecte TCGA-LUAD estratificada per l'estadi 4. Es representa la probabilitat de supervivència a l'eix de les "y" en funció del temps a l'eix de les "x". La corba negra correspon a l'estadi 1 del tumor, la corba vermella correspon a l'estadi 2 del tumor, la corba verda correspon a l'estadi 3 del tumor i la corba de color blau correspon a l'estadi 4 del tumor. S'afegeix el test estadístic per a comprovar si les diferències entre les corbes són significatives, en aquest cas no ho són perquè el p-valor és major que 0.05. Dins del projecte hi havia valors els quals no es va reportar quin era l'estadi del tumor del pacient. Aquests han sigut eliminats i no han sigut considerats per l'anàlisi.

Es volia verificar si la principal diferència entre la supervivència dels grups d'interès era deguda a la possible signatura de TGF- β o senescència i no a altres factors com l'estadi del tumor, l'edat dels pacients o el sexe. Per una banda, es va aplicar un model de Cox per determinar si les corbes de supervivència entre els tractaments presentaven diferències significatives segons l'estadi al qual es trobava el tumor (*taula 4*). Per altra banda, es va realitzar un segon model de Cox per establir si les corbes de supervivència entre els pacients tractats i els no tractats presentaven diferències significatives segons el sexe i l'edat d'aquests (*taula 5*).

En les taules que s'exposen a continuació, podem veure com tant el tractament aplicat com l'estadi són significatius; però en canvi ni l'edat ni el sexe ho són. Fet que indica que existeix una associació positiva entre el tractament aplicat als pacients i l'estadi en què es troben els seus tumors, però en canvi no existeix aquesta relació entre l'edat dels pacients ni el sexe d'aquests amb el tractament.

Taula 4. Resultat de la regressió pel model de Cox aplicada per tractament administrat i estadi del tumor dels pacients. Dades extretes del projecte TCGA-LUAD. La variable “clin_df3\$treat” és la que correspon al tractament administrat als pacients (quimioteràpia i sense), mentre que la variable “clin_df3\$stage” correspon als estadis en què es trobaven els tumors dels pacients. S’indica amb un requadre vermell la columna “p”. Aquesta variable correspon al p-valor de les corbes de supervivència i aquest es considera significatiu quan és menor a 0.05, ja que aquest model es resol com un model de regressió lineal.

| | coef | exp(coef) | se(coef) | z | p |
|-----------------|---------|-----------|----------|-------|----------|
| clin_df3\$treat | 0.49129 | 1.63443 | 0.16854 | 2.915 | 0.00356 |
| clin_df3\$stage | 0.45295 | 1.57294 | 0.07577 | 5.978 | 2.26e-09 |

Likelihood ratio test=49.18 on 2 df, p=2.091e-11
n= 460, number of events= 169
(62 observations deleted due to missingness)

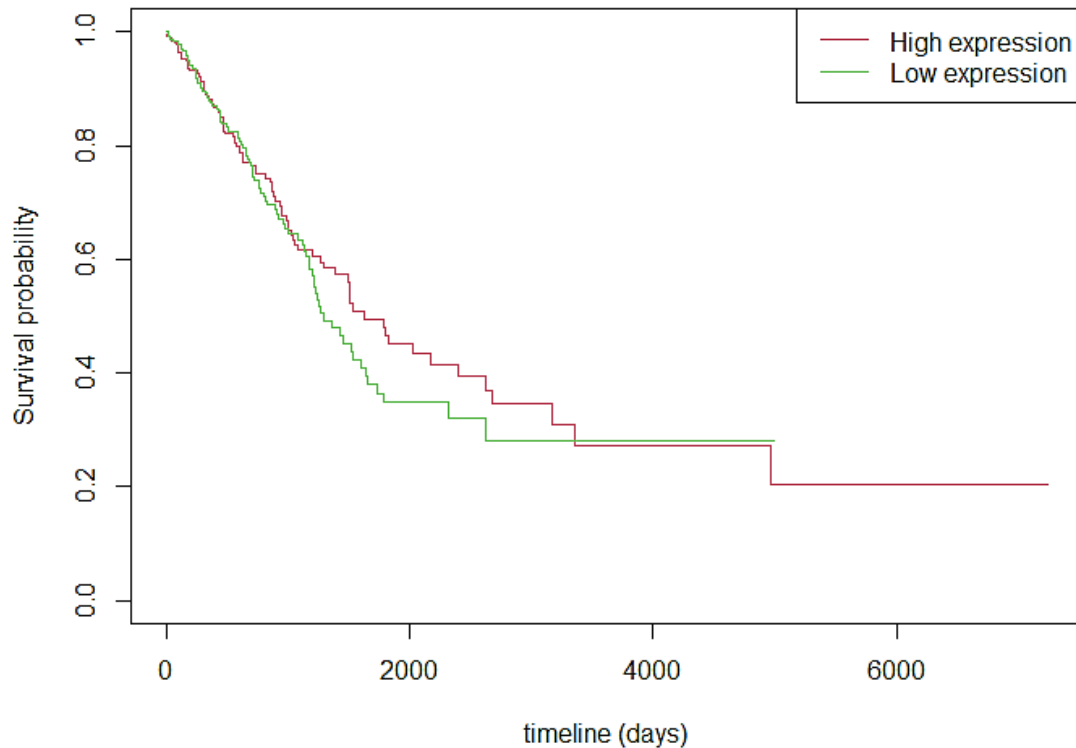
Taula 5. Resultat de la regressió pel model de Cox aplicada per tractament administrat, estadi del tumor, sexe i edat dels pacients. Dades extretes del projecte TCGA-LUAD. La variable “clin_df3\$treat” és la que correspon al tractament administrat als pacients (quimioteràpia i sense), mentre que la variable “clin_df3\$stage” correspon als estadis en què es trobaven els tumors dels pacients. La variable “clin_df3\$gendermale” és la que correspon al sexe dels pacients i finalment la variable “clin_df3\$age_at_index” correspon a l’edat d’aquests. S’indica amb un requadre vermell la columna “p”. Aquesta variable correspon al p-valor de les corbes de supervivència i aquest es considera significatiu quan és menor a 0.05, ja que aquest model es resol com un model de regressió lineal.

| | coef | exp(coef) | se(coef) | z | p |
|------------------------|-----------|-----------|----------|--------|----------|
| clin_df3\$treat | 0.506407 | 1.659318 | 0.169889 | 2.981 | 0.00287 |
| clin_df3\$stage | 0.438854 | 1.550929 | 0.077761 | 5.644 | 1.67e-08 |
| clin_df3\$gendermale | -0.060769 | 0.941041 | 0.159136 | -0.382 | 0.70256 |
| clin_df3\$age_at_index | 0.011359 | 1.011424 | 0.007865 | 1.444 | 0.14869 |

Likelihood ratio test=46.28 on 4 df, p=2.151e-09
n= 450, number of events= 166
(72 observations deleted due to missingness)

Un dels objectius principals era trobar una relació entre la possible signatura de senescència en tumors i la supervivència dels pacients, però després de realitzar l’anàlisi d’enriquiment dels gens (GSEA), es va veure que no hi havia presència de cap signatura de senescència. Tot i així, es va fer l’anàlisi de supervivència per a acabar de corroborar els resultats. En aquest, es van subdividir els pacients en dos grups: els que tenien alta expressió de marcadors de senescència i els que tenien baixa expressió, però tal com s’observa a la *figura 19*, no hi ha diferències significatives entre les dues corbes.

A favor de les nostres expectatives, els resultats no mostren evidències que la supervivència es vegi afectada per la signatura de senescència als pacients tractats amb quimioteràpia, per tant, no podem concloure cap relació entre la signatura de senescència en els tumors i la supervivència dels pacients.



n=391, 131 observations deleted due to missingness.

| | N | Observed | Expected | (O-E) ² /E | (O-E) ² /V |
|------------------------------|-----|----------|----------|-----------------------|-----------------------|
| data\$clinical_patient.sen=0 | 201 | 30 | 34.3 | 0.527 | 1.1 |
| data\$clinical_patient.sen=1 | 190 | 36 | 31.7 | 0.569 | 1.1 |

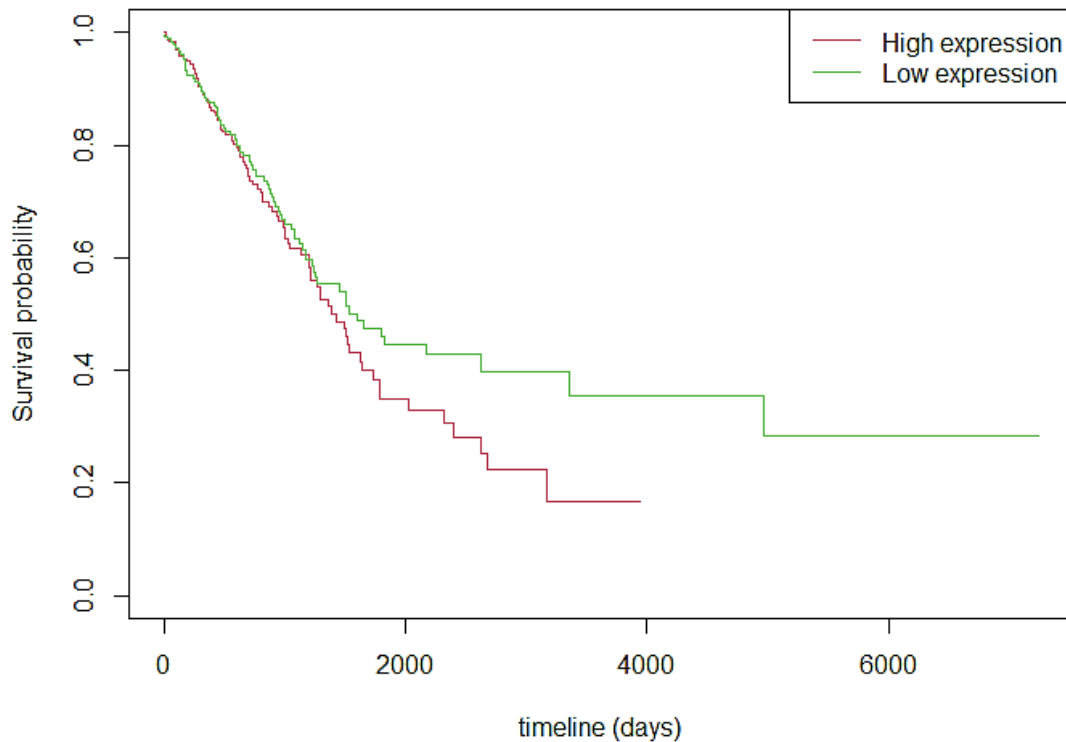
Chisq= 1.1 on 1 degrees of freedom, p= 0.3

Figura 19. Corba de supervivència Kaplan-Meier de pacients amb càncer de pulmó tractats amb quimioteràpia del projecte TCGA-LUAD segons els nivells d'expressió de marcadors de senescència. Es representa la probabilitat de supervivència a l'eix de les "y" en funció del temps a l'eix de les "x". La corba vermella correspon als pacients tractats amb quimioteràpia amb alta expressió de marcadors de senescència mentre que la verda correspon als pacients tractats amb quimioteràpia amb baixa expressió de marcadors de senescència. S'afegeix el test estadístic per a comprovar si les diferències entre les corbes són significatives, en aquest cas no ho són perquè el p-valor és major que 0.05. Dins del projecte hi havia valors els quals no es va reportar quin era el tractament que es va subministrar al pacient. Aquests han sigut eliminats i no han sigut considerats per l'anàlisi.

Com ja hem esmentat anteriorment, alguns dels efectes del TGF- β és la inducció de senescència, de manera que després de realitzar el GSEA per a aquesta via i veure que els gens diferencialment expressats es trobaven enriquits, es va proposar el mateix estudi per als seus marcadors. Així doncs, es volia trobar una relació entre la signatura de TGF- β en tumors i la supervivència dels pacients, subdividint-los en dos grups: pacients amb alta expressió de marcadors de TGF- β i pacients amb baixa expressió.

Al contrari de les nostres expectatives, els resultats mostren que la signatura de TGF- β en els tumors dels pacients tractats amb quimioteràpia no afecta la supervivència d'aquests (figura 20). El resultat estadístic denota que no hi ha diferències significatives entre les dues corbes, però s'observa que els pacients tractats amb alta expressió de marcadors de TGF- β tenen menys probabilitats de sobreviure.

De totes maneres, com ja hem mencionat, no podem concloure cap relació significativa entre la via de senyalització de TGF- β en els tumors i la supervivència dels pacients.



n=391, 131 observations deleted due to missingness.

| | N | Observed | Expected | (O-E) ² /E | (O-E) ² /V |
|------------------------------|-----|----------|----------|-----------------------|-----------------------|
| data\$clinical_patient.sen=0 | 182 | 36 | 30.1 | 1.16 | 2.13 |
| data\$clinical_patient.sen=1 | 209 | 30 | 35.9 | 0.97 | 2.13 |

Chisq= 2.1 on 1 degrees of freedom, p= 0.1

Figura 20. Corba de supervivència Kaplan-Meier de pacients amb càncer de pulmó tractats amb quimioteràpia del projecte TCGA-LUAD segons els nivells d'expressió de marcadors de TGF- β . Es representa la probabilitat de supervivència a l'eix de les "y" en funció del temps a l'eix de les "x". La corba vermella correspon als pacients tractats amb quimioteràpia amb alta expressió de marcadors de TGF- β , mentre que la verda correspon als pacients tractats amb quimioteràpia amb baixa expressió de marcadors de TGF- β . S'afegeix el test estadístic per a comprovar si les diferències entre les corbes són significatives, en aquest cas no ho són perquè el p-valor és major que 0.05. Dins del projecte hi havia valors els quals no es va reportar quin era el tractament que es va subministrar al pacient. Aquests han sigut eliminats i no han sigut considerats per l'anàlisi.

4.2. Estudi GEO65258

4.2.1. Control de qualitat

Un segon objectiu d'aquest estudi és determinar si en mostres de càncer de pulmó primerenques trobem evidència d'una signatura de senescència, ja que com s'ha introduït abans, la senescència també es pot induir en resposta a la sobreexpressió d'un oncogen durant la tumorigènesi. En aquest estudi, ens vam centrar en l'anàlisi de mostres de ratolí com s'explica en l'apartat 3.3. *Metodologia*. Per tal de procedir amb aquestes anàlisis i respondre les preguntes plantejades en els objectius, els primers passos van ser el control de qualitat per detectar *outliers* en les mostres, i eliminar-les en el cas que ho fossin. Gràcies al gràfic de densitats (*figura 21A*) i al diagrama de caixes d'intensitats (*figura 21B*) es va poder corroborar que les mostres eren de suficient qualitat per a no prescindir de cap d'elles.

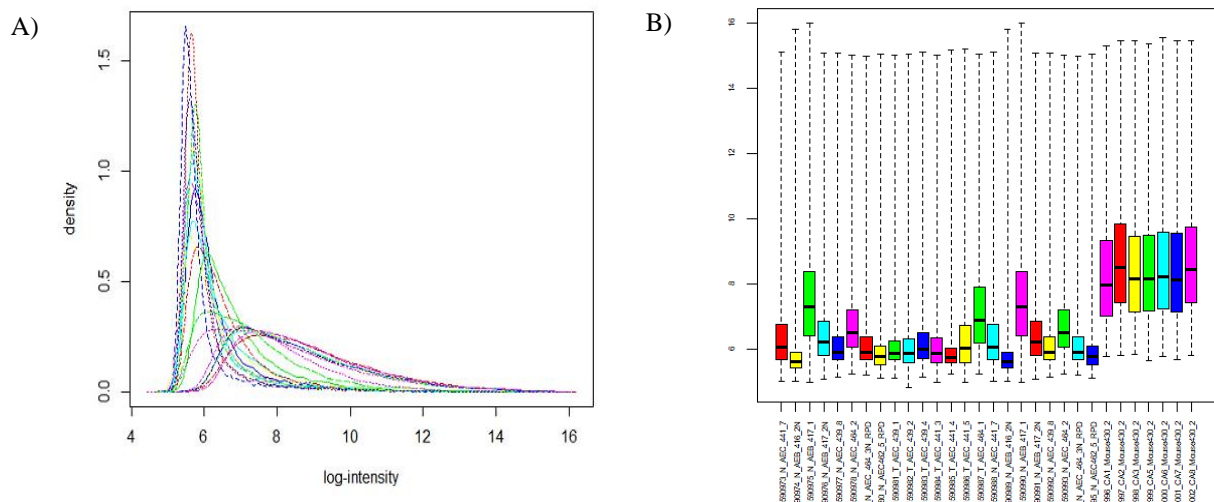


Figura 21. Resultats del control de qualitat de les dades de Microarray del projecte GSE65258. A) *Histograma:* Gràfic de densitats per a veure la distribució de les mostres. Cada color correspon a una mostra i es representa el logaritme de la intensitat respecte a la densitat mostral. B) *Diagrama de caixes:* Gràfic d'intensitat per a veure la distribució de les mostres d'acord amb la mediana. Cada caixa és una mostra i es representen les mostres en funció de la seva intensitat.

4.2.2. Agregació de les mostres

Per tal d'explorar l'agrupació de les mostres, es van generar uns clústers jeràrquics per a veure com aquestes es combinaven. Amb el fi d'obtenir un ventall de resultats més ampli, es van utilitzar diferents algorismes, però tal com es mostra a la *figura 22*, amb tots els mètodes de lincatge els clústers representats en dendrogrames s'agrupen per condició, és a dir, segons el tipus de lesió.

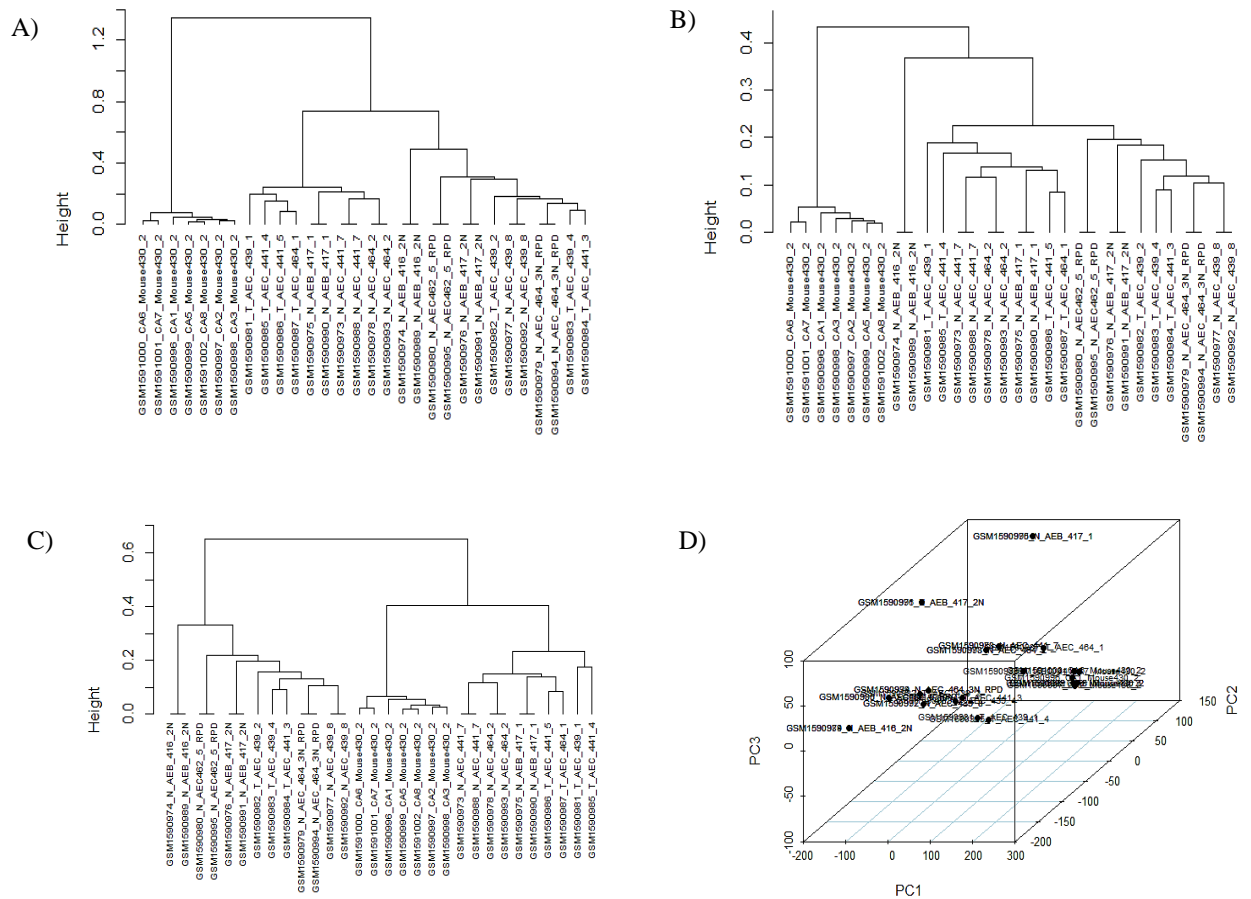


Figura 22. Clústers jeràrquics que representen l'agrupació de les mostres de les dades de Microarray del projecte GSE65258. A) Clúster jeràrquic representat en dendrogrames basat en el mètode 'ward2'. B) Clúster jeràrquic representat en dendrogrames basat en el mètode 'complete'. C) Clúster jeràrquic representat en dendrogrames basat en el mètode 'average'. D) Agrupació de les mostres per mitjà de la tècnica de la reducció de la dimensió 'PCA'.

4.2.3. Anàlisi d'expressió diferencial

Es van determinar els gens que s'expressaven diferencialment classificant les dades segons els dos grups d'estudi: els ratolins que tenien lesions hiperplàstiques (lesions primerenques) i els ratolins que tenien lesions de 'full-blown adenocarcinoma' (lesions avançades). Es van adquirir un total de 30 mostres, incloses setze mostres de la condició normal (control), set de la condició hiperplàstica i set de la condició 'full-blown adenocarcinoma'. Es van analitzar un total de 45.101 gens. Aquests es van filtrar posteriorment per un p-valor menor a 0.05 i un logaritme del 'Fold-Change' (*log fold-change, logFC*) major que 1, obtenint així, un total de 20.112 gens diferencialment expressats entre les dues condicions.

El 'volcano plot' (figura 23) i el 'heat map' (figura 24) mostren els nivells d'expressió d'aquests gens. En el 'volcano plot', s'hi pot observar més quantitat de gens amb p-valors infraexpressats estadísticament significatius en els ratolins amb lesions hiperplàstiques, els quals ronden els valors de logFC d'entre -5 i -1; mentre que els que es troben sobreexpressats en aquesta condició, romanen en un rang d'entre 1 i 4.

En el 'heat map', s'assenyalen les mostres de 'full-blown adenocarcinomes' amb un requadre vermell, i s'assenyalen les mostres hiperplàstiques amb un requadre blau.

Així doncs, entre aquests dos grups s'hi preveuen unes diferències d'expressió molt clares, ja que la majoria de gens que es troben sobreexpressats en lesions avançades es troben infraexpressats en lesions primerenques.

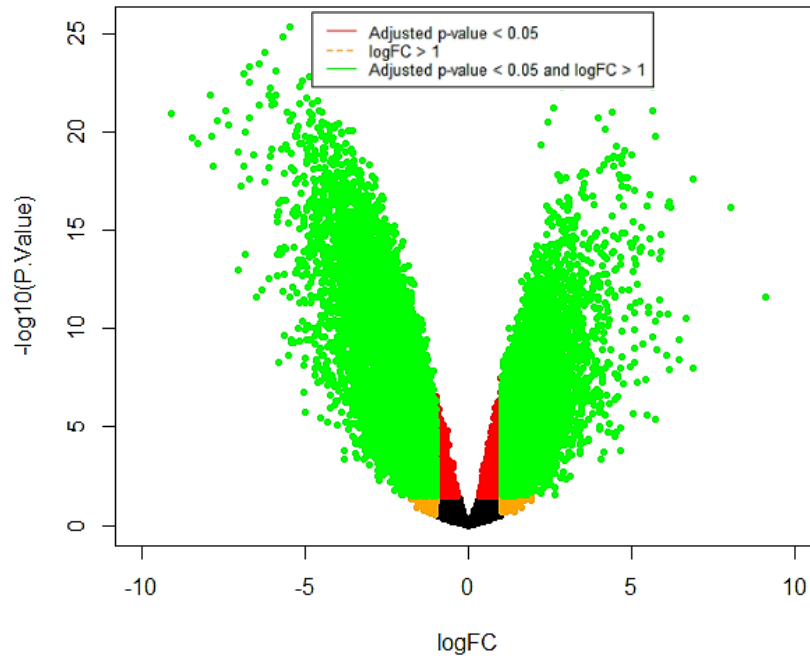


Figura 23. Volcano plot representant els gens diferencialment expressats del projecte GSE65258. Els punts vermells signifiquen aquells valors els quals el seu p-value és menor a 0.05. Els taronges són aquells els quals el seu log Fold-Change (logFC) és major a 1 i finalment, els verds són aquells als quals els corresponen les dues condicions.

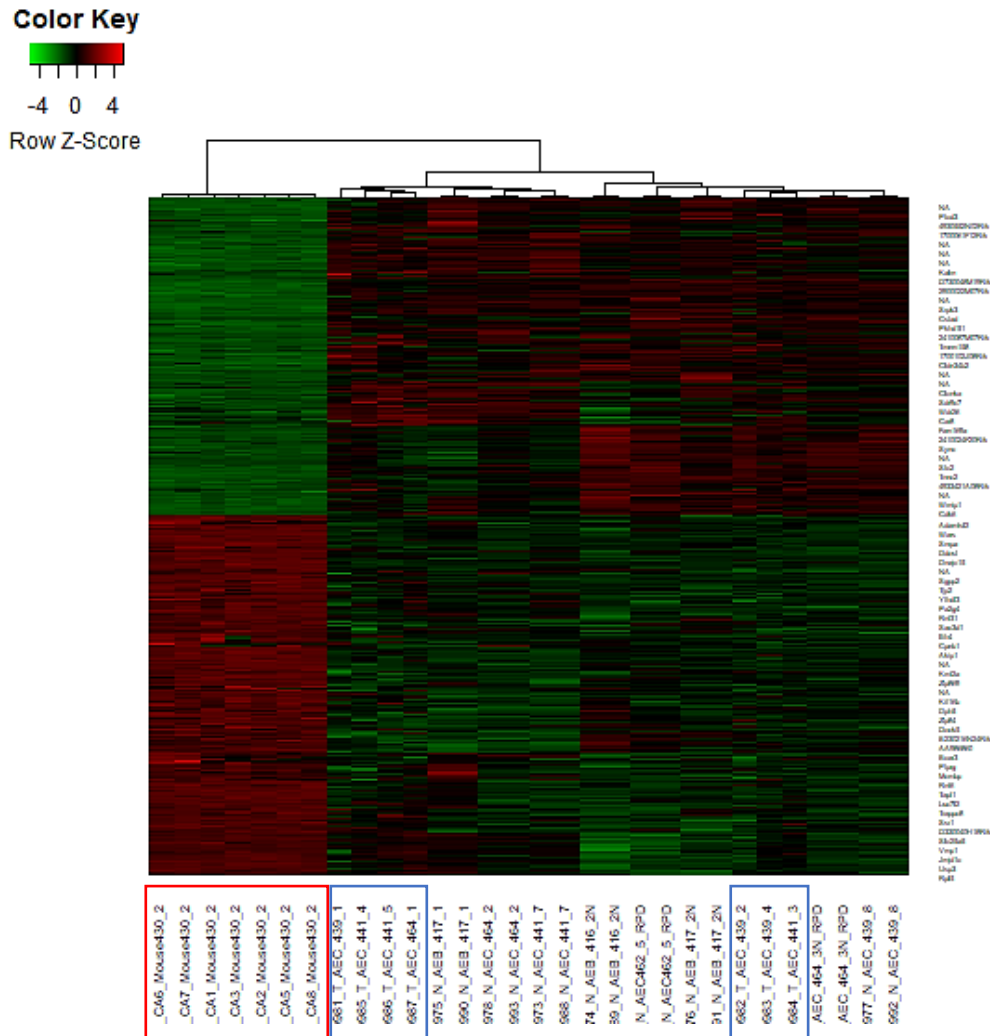


Figura 24. HeatMap representant els gens diferencialment expressats del projecte GSE65258. Les files representen els gens i les columnes representen les mostres. Els nivells d'expressió en aquest gràfic, tal com s'indica a la llegenda estan codificats per intensitat de color, essent el color vermell l'indicador d'alta expressió i el verd l'indicador de baixa expressió. Les mostres assenyalades amb un requadre vermell corresponen a les mostres de lesions avançades, mentre que les assenyalades amb el requadre blau corresponen a les mostres de lesions primerenques. Noteu que només s'ha anotat el nom dels primers 73 gens per tal que en el gràfic les diferències d'expressió fossin visibles però en el gràfic s'hi representen tots els gens diferencialment expressats. El nom de les mostres no es veu complet ja que aquests eren massa llargs.

4.2.4. Gene Set Enrichment Analysis (GSEA)

Per tal de determinar si els processos de senescència es troben enriquits en la progressió i el pronòstic del càncer de pulmó en *Mus Musculus*, es van analitzar el conjunt de gens considerats com a diferencialment expressats i es van comparar amb una signatura de senescència. Aquesta signatura conté 77 gens sobreexpressats en processos de senescència, i es va extreure de la pàgina oficial de GSEA (<https://www.gsea-msigdb.org/gsea>) realitzada per Fridman, AL i Tainsky, MA.

Es va interpretar el resultat de l'anàlisi amb un gràfic convencional que mostra els mil primers gens a l'eix de les 'x' (ordenats en successió), i el seu p-valor corresponent a l'eix de les 'y' (figura 25).

L'indicador que defineix si els gens estan enriquits és el p-valor, ja que com més petit és aquest més significatiu són els resultats. En altres paraules, perquè els gens es trobessin enriquits, els p-valors més petits i per tant més significatius, haurien d'estar al capdavant del gràfic. Però tal com s'observa a la figura, els gens diferencialment expressats no es troben enriquits en processos de senescència, ja que pels primers 500 gens, els p-valors respectius estan al voltant d'1.

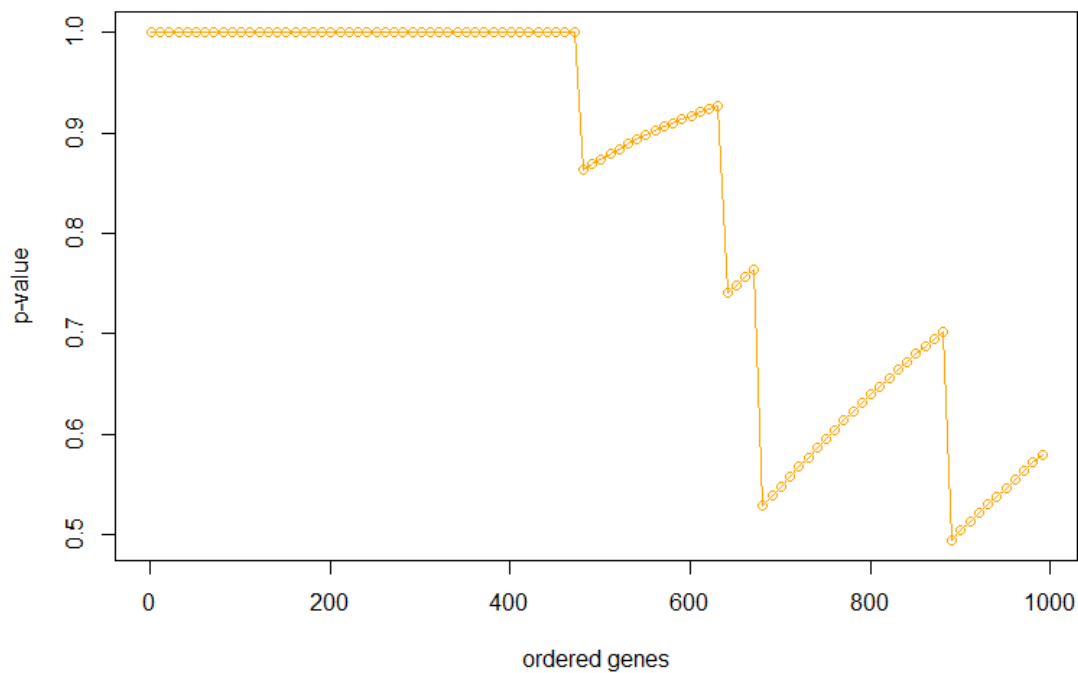


Figura 25. Resultats de l'anàlisi d'enriquiment de GSEA per una signatura de senescència en ratolins amb càncer de pulmó del projecte GSE65258. Com es pot observar al gràfic, els gens es troben representats en una successió de dos-cents en dos-cents. El seu p-valor respectiu es troba representat a l'eix de les "y", de manera que com més petit sigui aquest, més significatiu és el resultat.

5. Conclusions

L'adenocarcinoma de pulmó és la neoplàsia de pulmó més comuna, però malauradament molts mecanismes moleculars del seu desenvolupament i les seves causes no s'han aconseguit desxifrar de moment. Aquest tipus de càncer té una supervivència de menys del 10% durant els primers cinc anys després del diagnòstic. En aquest sentit, hi ha dos factors que tenen vital importància: el diagnòstic tardà i la ineficiència del tractament.

El primer d'aquests factors, el diagnòstic tardà, es deu a una manca de coneixement sobre els processos que ocorren a les lesions primerenques durant el desenvolupament d'adenocarcinoma. En conseqüència, la major part dels casos es detecten en estadis molt avançats, la qual cosa té un impacte molt important en la baixa supervivència.

El segon factor es deu al tractament actual per aquest tipus de càncer. Tot i que s'han fet alguns avenços, el tractament principal continua sent el mateix des de fa més de trenta anys: la quimioteràpia combinatòria amb platí. Tot i que els pacients reben un tractament, la gran majoria acaben desenvolupant una recaiguda en menys de dos anys, la qual cosa contribueix molt significativament a l'alta taxa de mortalitat. Això també indica un possible problema amb la quimioteràpia basada en platí.

Molts estudis es basen en l'anàlisi d'un únic procés o inclouen molt poques mostres rellevants, la qual cosa no sempre proporciona una explicació detallada dels mecanismes que hi poden haver durant el desenvolupament o tractament de càncer de pulmó. En canvi, una anàlisi global incloent anàlisis de transcriptomes, proteomes i genomes, gràcies a l'ús de les bases de dades, podrien ajudar a descriure'ls, i inclús, revelar-ne marcadors específics per la seva detecció i pronòstic.

Partint de les observacions que vam obtenir al laboratori del Dr. Muñoz-Espín, l'objectiu d'aquest estudi va ser analitzar dades clíniques de centenars de mostres recollides en bases de dades, per determinar la possible relació entre processos de senescència i aquests dos factors en càncer de pulmó: la quimioteràpia i el desenvolupament de lesions canceroses primerenques.

Després d'haver analitzat les dades clíniques del projecte *TCGA-LUAD*, es confirma que els pacients amb adenocarcinoma de pulmó que reben quimioteràpia presenten una taxa de supervivència menor que els pacients que no la reben. Tot i que aquests resultats van en línia de la nostra hipòtesi, cal tenir en compte que hi ha altres factors clínics que poden contribuir a aquesta relació. Per exemple, els pacients que es tracten amb quimioteràpia generalment presenten estadis més avançats, i tot i que reben tractament farmacològic, això també pot afectar la supervivència.

Tenint en compte aquest aspecte, es van realitzar unes anàlisis de supervivència estratificades en funció de l'estadi en què es trobaven els tumors dels dos grups d'interès (pacients tractats i pacients no tractats). En estadis entremitjos (estadis 2 i 3) no hi ha pràcticament diferències en la supervivència dels pacients, sobretot en els primers mesos. En canvi, en estadis primerencs (estadi 1), els pacients que han rebut un tractament farmacològic tenen menys supervivència, mentre que en estadis avançats (estadi 4), els pacients que s'han sotmès a

quimioteràpia tenen més supervivència en comparació amb els que no es van sotmetre a cap tractament. Tot i que les diferències en la corba de supervivència de l'estadi 4 no fossin significatives, aquestes evidències ens podrien indicar que la quimioteràpia podria ser efectiva només per estadis avançats i que en canvi per estadis primerencs aquesta teràpia podria ser contraproduent.

Gràcies a les anàlisis estadístiques realitzades a partir de la regressió pel mètode de *Cox* es va determinar una associació significativa entre la supervivència i l'estadi en què es trobaven els tumors dels pacients. En canvi, no es va determinar aquesta relació per l'edat ni el sexe dels pacients.

Havent obtingut evidències que la quimioteràpia està lligada a una pitjor supervivència, es van analitzar les pautes d'expressió gènica dels tumors tractats i no tractats.

Els resultats de les anàlisis d'expressió diferencial indiquen que hi ha una gran quantitat de gens sobreexpressats en pacients tractats amb quimioteràpia en comparació amb els perfils d'expressió dels pacients no tractats. Les anàlisis d'enriquiment per GO i KEGG demostren un llistat de processos enriquits en aquests pacients. Tot i això, al realitzar l'anàlisi d'enriquiment per GSEA no es va poder demostrar que els processos de senescència es trobessin enriquits entre els pacients tractats. Per aquest motiu, no es van trobar resultats significatius en un dels objectius d'aquesta anàlisi, que era trobar una relació entre la signatura de senescència dels tumors tractats i la supervivència dels pacients. En conclusió, aquests resultats rebutgen la hipòtesi formulada a l'inici d'aquest estudi: "els pacients que reben quimioteràpia tenen major expressió de marcadors de senescència en els seus tumors i, per conseqüència, una pitjor prognosi (taxa de supervivència disminuïda)."

Paral·lelament, amb l'objectiu d'entendre la possible relació entre la senescència i el desenvolupament d'adenoma, es va fer l'anàlisi de les pautes d'expressió gènica de mostres de *Mus musculus* de lesions primerenques o 'hiperplàstiques' comparat amb lesions de tumors avançats o 'full-blown adenocarcinomes'.

Les evidències de les anàlisis d'expressió diferencial indiquen que hi ha una major presència de gens infraexpressats en lesions hiperplàstiques en comparació amb lesions avançades o de 'full-blown adenocarcinoma', i a més, s'hi preveuen grans diferències d'expressió entre les dues condicions. En fer l'anàlisi per GSEA de la signatura de senescència, però, es confirma que les lesions primerenques no expressen marcadors de senescència.

Resumint, durant aquest estudi s'ha fet una anàlisi bioinformàtica exhaustiva del transcriptoma de mostres clíniques humanes i murines, amb l'objectiu de detectar possibles associacions entre: (1) senescència, quimioteràpia i supervivència, i (2) senescència i lesions primerenques en càncer de pulmó. En conclusió, s'ha demostrat que, tot i que hi ha diferències significatives en l'àmbit d'expressió en molts processos entre els grups analitzats, no hi ha evidència d'una signatura de senescència, rebutjant les nostres hipòtesis inicials.

En un futur, es podria intentar focalitzar l'anàlisi a grups determinats de pacients, com ara aquells que presenten un mateix estadi, o ampliar el nombre de mostres amb l'ús d'altres bases de dades potencialment útils.

6. Bibliografia

- Braig, M. *et al.* (2005) ‘Oncogene-induced senescence as an initial barrier in lymphoma development’, *Nature*, 436(7051), pp. 660–665. doi: 10.1038/nature03841.
- Campisi, J. (2005) ‘Senescent cells, tumor suppression, and organismal aging: Good citizens, bad neighbors’, *Cell*. Cell Press, pp. 513–522. doi: 10.1016/j.cell.2005.02.003.
- Campo-Trapero, J. *et al.* (2008) ‘Cellular senescence in oral cancer and precancer and treatment implications: A review’, *Acta Oncologica*. Taylor & Francis, pp. 1464–1474. doi: 10.1080/02841860802183612.
- Casado-Vela, J. *et al.* (2011) ‘Approaches for the study of cancer: Towards the integration of genomics, proteomics and metabolomics’, *Clinical and Translational Oncology*, 13(9), pp. 617–628. doi: 10.1007/s12094-011-0707-9.
- Collado, M. and Serrano, M. (2010) ‘Senescence in tumours: Evidence from mice and humans’, *Nature Reviews Cancer*. Europe PMC Funders, pp. 51–57. doi: 10.1038/nrc2772.
- Corral, L. B. (2016) *Análisis de perfiles de expresión génica*. Available at: <http://hdl.handle.net/10481/44640> (Accessed: 22 April 2020).
- Cotarelo Fernández, A. (2016) *Marcadores en cáncer: estrategias de análisis de expresión múltiple*.
- Dela Cruz, C. S., Tanoue, L. T. and Matthay, R. A. (2011) ‘Lung Cancer: Epidemiology, Etiology, and Prevention’, *Clin Chest Med*, 32(4). doi: 10.1016/j.ccm.2011.09.001.
- Dasari, S. and Bernard Tchounwou, P. (2014) ‘Cisplatin in cancer therapy: Molecular mechanisms of action’, *European Journal of Pharmacology*. Elsevier, pp. 364–378. doi: 10.1016/j.ejphar.2014.07.025.
- Doll, R. *et al.* (2004) ‘Mortality in relation to smoking: 50 Years’ observations on male British doctors’, *British Medical Journal*. British Medical Journal Publishing Group, 328(7455), pp. 1519–1528. doi: 10.1136/bmj.38142.554479.ae.
- Genes, ADN y cáncer | Cancer Research UK* (2020). Available at: <https://www.cancerresearchuk.org/about-cancer/what-is-cancer/genes-dna-and-cancer> (Accessed: 24 April 2020).
- Guancial, E. A. *et al.* (2016) ‘Platinum concentration and pathologic response to cisplatin-based neoadjuvant chemotherapy in muscle-invasive bladder cancer’, *PLoS ONE*. Public Library of Science, 11(5), p. e0155503. doi: 10.1371/journal.pone.0155503.
- Haas, B. J. and Zody, M. C. (2010) ‘Advancing RNA-Seq analysis’, *Nature Biotechnology*. Nature Publishing Group, pp. 421–423. doi: 10.1038/nbt0510-421.
- Hanahan, D. and Weinberg, R. A. (2011) ‘Hallmarks of cancer: The next generation’, *Cell*, pp. 646–674. doi: 10.1016/j.cell.2011.02.013.
- Identification of candidate processes and mechanisms involved in lung* (2020). Available at: <https://www.munoz-espinlab.com/identification-of-candidate-process> (Accessed: 27 April 2020).

Lecot, P. *et al.* (2016) 'Context-dependent effects of cellular senescence in cancer development', *British Journal of Cancer*. Nature Publishing Group, pp. 1180–1184. doi: 10.1038/bjc.2016.115.

Lung cancer - NHS (2020). Available at: <https://www.nhs.uk/conditions/lung-cancer/> (Accessed: 25 April 2020).

Lung cancer - Treatment - NHS (2020). Available at: <https://www.nhs.uk/conditions/lung-cancer/treatment/> (Accessed: 26 April 2020).

M. Parra (2018) *CARACTERIZACIÓN TRANSCRIPTÓMICA DEL CÁNCER DE ENDOMETRIO Y SU IMPORTANCIA EN EL DISEÑO DE NUEVAS ESTRATEGIAS TERAPÉUTICAS*.

Marín, M. (2014) *Uso de datos genómicos para la búsqueda de reguladores de importancia en cáncer*.

Muñoz-Espín, D. *et al.* (2013) 'XProgrammed cell senescence during mammalian embryonic development', *Cell*. Cell Press, 155(5), p. 1104. doi: 10.1016/j.cell.2013.10.019.

NCBI - National Center For Biotechnology Information (2020) *Gene Expression*. Available at: <https://www.ncbi.nlm.nih.gov/probe/docs/applexpression/> (Accessed: 3 May 2020).

Roberson, R. S. *et al.* (2005) 'Escape from therapy-induced accelerated cellular senescence in p53-null lung cancer cells and in human lung cancers', *Cancer Research*. American Association for Cancer Research, 65(7), pp. 2795–2803. doi: 10.1158/0008-5472.CAN-04-1270.

Sager, M. *et al.* (2015) 'Transcriptomics in cancer diagnostics: Developments in technology, clinical research and commercialization', *Expert Review of Molecular Diagnostics*. Taylor and Francis Ltd, pp. 1589–1603. doi: 10.1586/14737159.2015.1105133.

Sebestyén, E. *et al.* (2016) 'Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks', *Genome Research*. Cold Spring Harbor Laboratory Press, 26(6), pp. 732–744. doi: 10.1101/gr.199935.115.

Transcriptomics - Latest research and news | Nature (2020). Available at: <https://www.nature.com/subjects/transcriptomics> (Accessed: 3 May 2020).

Types of cancer | Cancer Research UK (2020). Available at: <https://www.cancerresearchuk.org/what-is-cancer/how-cancer-starts/types-of-cancer> (Accessed: 24 April 2020).

Wang, Q. *et al.* (2013) 'Polyploidy road to therapy-induced cellular senescence and escape', *International Journal of Cancer*, 132(7), pp. 1505–1515. doi: 10.1002/ijc.27810.

What Is Cancer? - National Cancer Institute (2020). Available at: <https://www.cancer.gov/about-cancer/understanding/what-is-cancer> (Accessed: 24 April 2020).

What is cancer? | Cancer Research UK (2020). Available at: <https://www.cancerresearchuk.org/about-cancer/what-is-cancer> (Accessed: 24 April 2020).

Yazmín González-Puertos, V. *et al.* (2015) *Participación del fenotipo secretor de las células senescentes en el desarrollo del cáncer, el envejecimiento y las enfermedades asociadas a la edad gaceta médica de México artículo de revisión Correspondencia*. Available at: www.anmm.org.mx

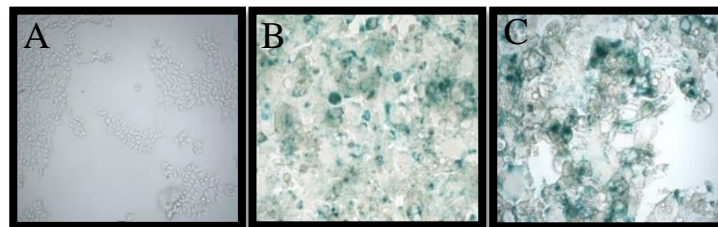
Annexes

A.1. Evidències experimentals obtingudes durant l'estada de pràctiques a la Universitat de Cambridge

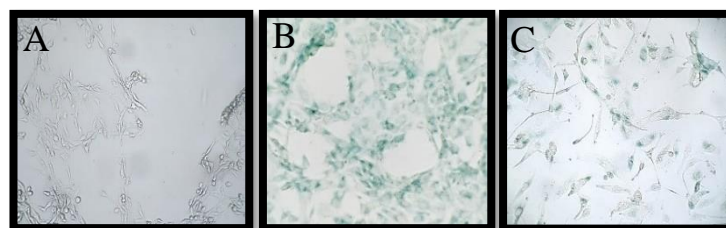
- Activitat β -Galactosidasa

Les cèl·lules van ser tractes amb fàrmacs de quimioteràpia durant set dies. Es van rentar amb PBS i es van fixar amb 4% paraformaldehid (PFA). Posteriorment es van tenyir per l'activitat β -Galactosidasa associada a senescència, emprant el “*Senescence β -Galactosidase Staining kit (Cell Signaling Technologies)* – en anglès –” seguint les instruccions del fabricant.

Les cèl·lules es van incubar durant entre sis i vuit hores a 37°C amb la solució de tinció que conté X-Gal, en dimetilformamida N-N després d'ajustar el pH a 6,0. Després de la tinció, les cèl·lules es van tornar a rentar per tal de prendre les imatges de les *imatges 1 i 2* amb el microscopi modular *Olympus Compact Brightfield (Life technologies)* a una ampliació de 10X.



Imatge 1. Imatges a magnificació 10X de les cèl·lules CML75 tenyides per l'activitat β -Galactosidasa després de 7 dies sota tractament. A) Cèl·lules control (sense tractament) B) Cèl·lules tractades amb Cisplatí a 3 μ M C) Cèl·lules tractes amb Docetaxel a 20nM. El color blau indica expressió de l'enzim β -Galactosidasa mentre que el color blanc a les colònies indica que no hi ha expressió de tal enzim. Així doncs, només aquelles cèl·lules induïdes a senescència per mitjà de components quimioterapèutics tindran presència de l'enzim β -Galactosidasa.



Imatge 2. Imatges a magnificació 10X de les cèl·lules MLg tenyides per l'activitat β -Galactosidasa després de 7 dies sota tractament. A) Cèl·lules control (sense tractament) B) Cèl·lules tractades amb Cisplatí a 5 μ M C) Cèl·lules tractes amb Docetaxel a 150nM. El color blau indica expressió de l'enzim β -Galactosidasa mentre que el color blanc a les colònies indica que no hi ha expressió de tal enzim. Així doncs, només aquelles cèl·lules induïdes a senescència per mitjà de components quimioterapèutics tindran presència de l'enzim β -Galactosidasa.

- SDS-PAGE i Western Blotting

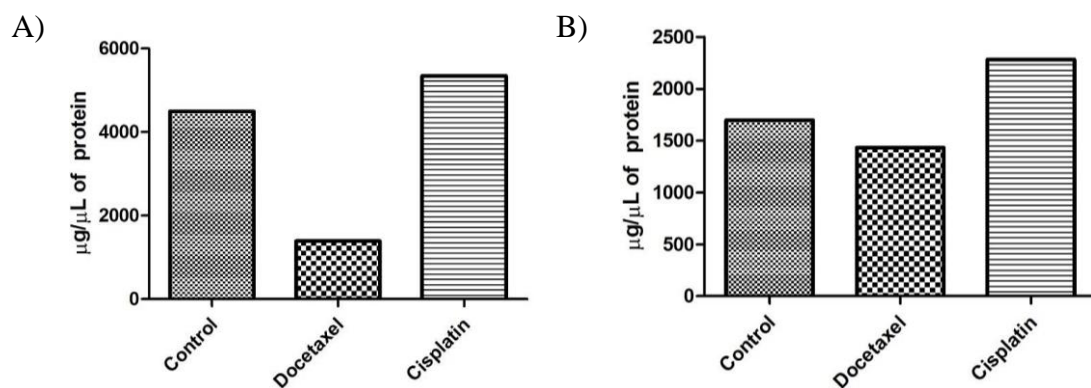
➤ *Producció del lisat i realització de l'electroforesi*

Es va procedir a la lisi de les cèl·lules amb 350 μ L del buffer RIPA – *Radioimmunoprecipitation Assay* – en anglès. Aquest buffer que està constituït per desoxicolat de

sodi, Tritan-X100, Tris, SDS i Clorur de sodi, es va suplementar amb EDTA i amb un còctel de inhibidors de proteases, per tal d'així bloquejar l'activitat de les proteases endògenes.

Es van raspar les cèl·lules utilitzant un aixecador de polietilè per a així transferir-les a un tub de centrifugació i incubar-les en gel durant 15 minuts. Posteriorment es van centrifugar els lisats a 4°C i a 15.000g durant 15 minuts i es va recollir el sobrenedant per transferir les proteïnes a un nou Eppendorf per a l'anàlisi de Western Blot.

Es va quantificar el lisat proteic utilitzant el kit “PierceTM BCA Protein Assay Kit”. Es va preparar una corba estàndard de BSA amb un rang de concentració final entre 25µg/mL i 2000µg/mL. Les mostres de proteïnes es van mesclar amb la solució seguint un rang consecutiu entre 1:2-4 i es van dipositar a una placa de 96-pous de “Thermo ScientificTM”. Aquesta placa es va incubar durant trenta minuts a 37°C i posteriorment es va analitzar per tal de quantificar les mostres (tal com es mostra a la imatge 3) i poder realitzar el gel d'electroforesis.



Imatge 3. Concentració de proteïnes present a les cèl·lules un cop es van llegir les absorbàncies del lisat proteic realitzat amb el “PierceTM BCA Protein Assay Kit”. A) Línia cel·lular CML75 B) Línia cel·lular MLg. Es pot observar per cada condició aplicada a les cèl·lules quina concentració de proteïna es va obtenir. Les que van estar sota tractament amb Cisplatí són les que tenen concentracions més elevades. Les tractades amb Docetaxel són les que tenen una concentració més baixa.

Es van diluir un total de 30µg de proteïna de cada mostra amb un buffer per carregar-ho a un gel preparat (*Mini-PROTEAN® TGXTM Precast Gel*). Aquest es va resoldre per uns 45 minuts a un voltatge constant de 200V immers en *SDS Running Buffer*. Les proteïnes es van electrotransferir del gel a una membrana de difluorur de polivinilidè (PVDF), essent l'última prèviament activada durant 60 segons en una solució de Metanol al 100%. El gel es va col·locar a un tanc de transferència electroforètica contingut en un buffer de transferència. Es va resoldre a 4°C durant tota la nit a un voltatge de 10V.

➤ Immunoblotting

Es va netejar la membrana amb Tris Buffered Saline complementat amb Tween 20 (TBS-T) i posteriorment es va incubar en una solució de TBS-T amb 5% de llet durant 1h a temperatura ambient. Les proteïnes presents a la llet s'uneixen a la membrana, minimitzant les unions inespecífiques dels anticossos primaris.

Per extreure l'excés de la solució bloquejant, la membrana es va rentar diverses vegades amb TBS-T durant 10 minuts cada una.

Tant els anticossos primaris com els secundaris es van diluir en 5% de llet en les quantitats que s'indiquen a la *imatge 4* i a la *imatge 5*, respectivament.

Posteriorment, la membrana es va incubar amb els anticossos secundaris durant tota la nit a 4°C en continu moviment. Es va procedir a realitzar el mateix protocol que per la incubació dels anticossos primaris, es va rentar la membrana tres vegades amb la solució de TBS-T per retirar els excessos.

| Anticòs | Espècie | Dilució | Proveïdor |
|----------------------|----------------|----------------|-----------------------------|
| Anti-GADPH | Rabbit | 1:2000 | ABCAM |
| Anti-Rb | Mouse | 1:2000 | Cell Signaling |
| Anti-Phospho-Rb | Rabbit | 1:1000 | Cell Signaling |
| Anti-p53 | Mouse | 1:500 | Santa Cruz Biotechnology |
| Anti-phospho-p53 | Rabbit | 1:1000 | Cell Signaling |
| Anti-p21 | Goat | 1:250 | Santa Cruz Biotechnology |
| Anti-CDKN2A/p16INK4a | Rabbit | 1:2000 | Abcam |

Imatge 4. Llistat d'anticossos primaris. També s'esmenten les condicions utilitzades per a la realització del Western Blot.

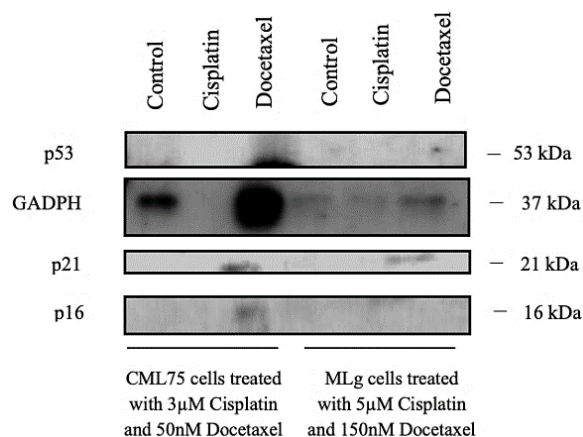
| Anticòs | Espècie | Dilució | Proveïdor |
|--|----------------|----------------|---------------------------|
| HRP-conjugated AffiniPute Anti-Rabbit IgG | Donkey | 1:5000 | Jackson ImmunoResearch |
| HRP-conjugated AffiniPute Anti-Mouse IgG | Donkey | 1:5000 | Jackson ImmunoResearch |
| HRP-conjugated AffiniPute Anti-Goat IgG | Donkey | 1:5000 | Jackson ImmunoResearch |

Imatge 5. Llistat d'anticossos secundaris. També s'esmenten les condicions utilitzades per a la realització del Western Blot.

La membrana es va incubar amb una solució de quimioluminiscència durant 1 minut a la foscor, per tal d'així posar la membrana en un film de raigs X i resoldre-ho utilitzant un Xograph Compact X4 automatic processor tal com es mostra a la *imatge 7*.

| Proteïna | Pes en kDa |
|---|------------|
| Rb – RB1 – Retinoblastoma protein | 110 |
| p53 – TRP53 – phosphoprotein 53 – Transformation Relation protein 53 | 53 |
| GADPH – Glyceraldehyde 3 phosphate | 37 |
| p21 – p21Cip1 – Cyclin Dependent Kinase inhibitor 1 | 21 |
| p16 - p16 ^{INK4a} – Cyclin Dependent inhibitor 2A | 16 |
| P- Rb – RB1 – Retinoblastoma protein phosphorylated | 110 |
| P-p53 – TRP53 – phosphoprotein 53 – Transformation Relation protein 53 phosphorylated | 53 |
| P- p21 – p21Cip1 – Cyclin Dependent Kinase inhibitor 1 | 21 |
| P- p16 - p16 ^{INK4a} – Cyclin Dependent kinase inhibitor 2A phopshorylated | 16 |

Imatge 6. Llistat de proteïnes detectables pel Western Blot. També s'esmenten els corresponents pesos moleculars per tal de detectar-ho a la membrana.



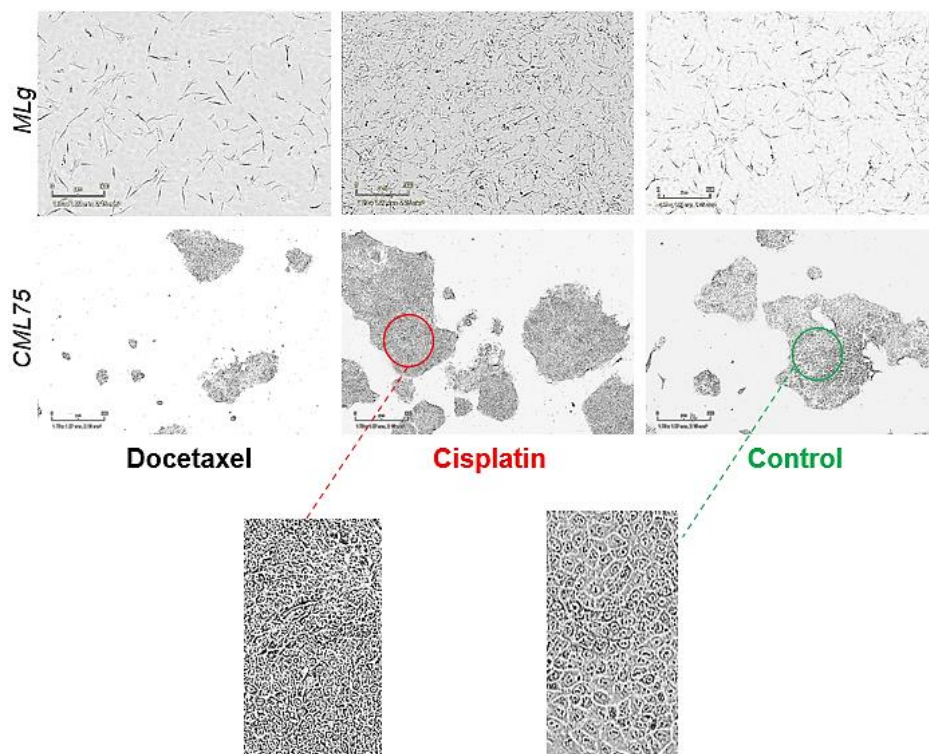
Imatge 7. Imatge representativa de l'anàlisi per Western Blot. S'hi veuen representades les dues línies cel·lulars i la seva condició respectiva. Aquestes cèl·lules van ser testades per les proteïnes pRB, Rb, Pp53, p53, p21 i p16. El GADPH es va emprar com a control de càrrega. Per tal d'analitzar la membrana, aquesta es va tallar en tres parts, cada una corresponent a dos parells de proteïnes que es detecten al mateix rang com es pot veure a la figura.

- Assajos de proliferació
- *Pel mètode convencional*

Es vol avaluar l'efecte que té el SASP a escala paracrina en la proliferació de les cèl·lules, és per això que es van deixar sota tractament quimioterapèutic durant set dies. Passat aquest temps, es van rentar amb PBS i es van suplementar amb medi fresc durant 48 hores.

Es va recol·lectar el medi amb el qual estaven aquestes cèl·lules i es va centrifugar un primer cop durant 5 minuts a 1240rpm, i un segon cop durant 10 minuts a 2500rpm.

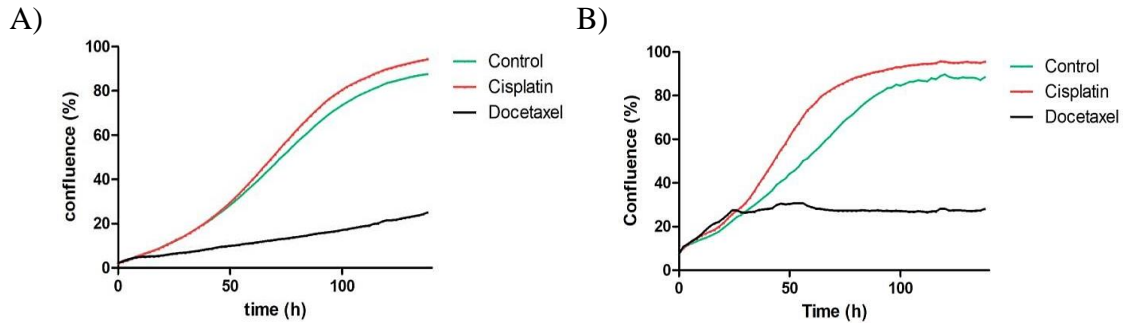
Aquest medi va ser afegit a unes cèl·lules que prèviament s'havien sembrat amb una confluència del 10% en un plat de 24 pous. Es van deixar créixer durant un període d'entre 72 i 112 hores i l'*IncuCyte*® capturava imatges com es pot veure a la *imatge 8*.



Imatge 8. Efecte del SASP induït per quimioteràpia a les dues línies cel·lulars. S'hi veuen representades cada una de les condicions de cada línia cel·lular amb un augment especial en les condicions de cisplatí i control per a poder fer una comparació més acurada, ja que les CML75 creixen apilades.

Es va recol·lectar el medi amb el qual estaven aquestes cèl·lules i es va centrifugar un primer cop durant 5 minuts a 1240rpm, i un segon cop durant 10 minuts a 2500rpm.

Gràcies al mateix software que l'*IncuCyte*® proporciona es van poder extreure dades de confluència en funció del temps que van ser representades a la *imatge 9* per cada línia cel·lular.



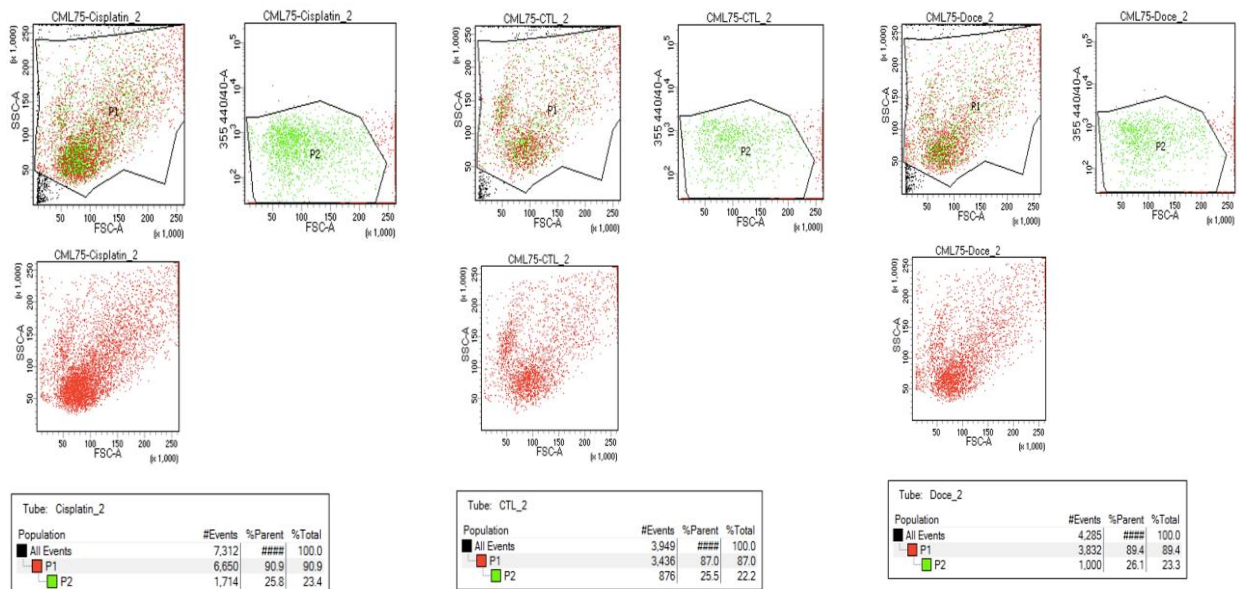
Imatge 9. Confluència cel·lular en percentatge de la línia cel·lular en funció del temps. A) Línia cel·lular CML75 B) Línia cel·lular MLG. Cada condició està representada amb un color i cada punt representa la confluència cel·lular en un temps determinat. Per cada repetició biològica es van realitzar tres pous per cada condició on es van capturar 9 imatges cobrint tot el pou.

➤ Per Flow Cytometry

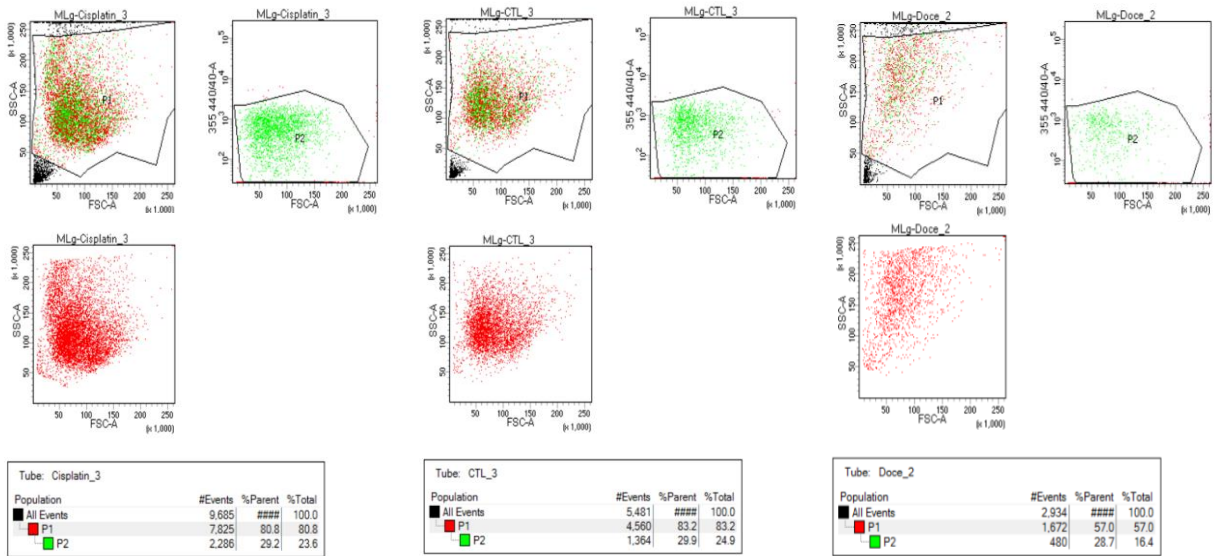
Es vol avaluar l'efecte que té el SASP a escala paracrina en la proliferació de les cèl·lules, és per això que es van deixar sota tractament quimioterapèutic durant set dies. Passat aquest temps, es van rentar amb PBS i es van suplementar amb medi fresc durant 48 hores.

Es va recol·lectar el medi amb qual estaven aquestes cèl·lules i es va centrifugar un primer cop durant 5 minuts a 1240rpm, i un segon cop durant 10 minuts a 2500rpm.

Aquest medi va ser afegit a unes cèl·lules que prèviament s'havien sembrat amb una confluència del 10% en un plat de 24 pous. Es van deixar créixer durant un període d'entre 72 i 112 hores i per a tripsinitzar-les i recol·lectar-les en un tub de polietilè per a poder quantificar amb la instrumentària del Flow Cytometer (*imatges 10 i 11*).

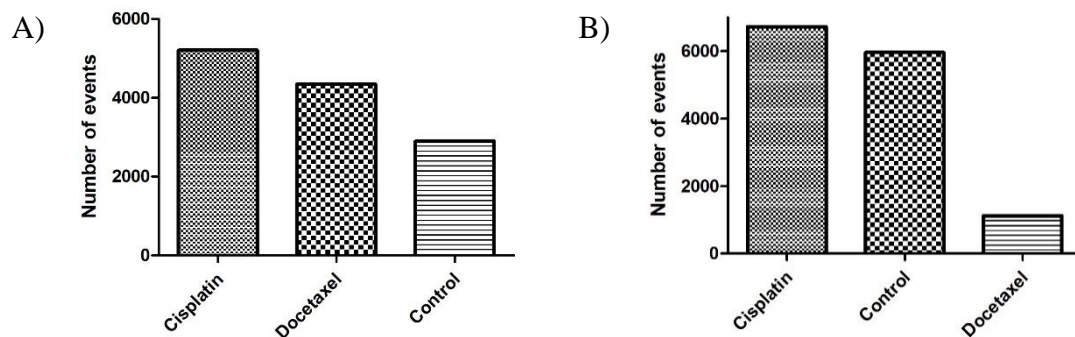


Imatge 10. Confluència cel·lular de la línia cel·lular CML75 obtinguda mitjançant la tècnica "Flow Cytometry". S'hi representen totes les condicions i cada una d'elles té tres gràfics que corresponen a diferents poblacions i una taula on apareix la confluència de cada població. Les cèl·lules tractades amb Cisplatin són les que tenen una major confluència, i les que tenen menor confluència són les cèl·lules control.



Imatge 11. Confluència cel·lular de la línia cel·lular MLg obtinguda mitjançant la tècnica “Flow Cytometry”. S’hi representen totes les condicions i cada una d’elles té tres gràfics que corresponen a diferents poblacions i una taula on apareix la confluència de cada població. Les cèl·lules tractades amb Cisplatí són les que tenen una major confluència, seguides del control i les que van ser tractades amb Docetaxel són les que tenen una menor confluència cel·lular.

A més, gràcies a les dades que el mateix software del Flow Cytometer ens va proporcionar, es va poder representar el total de cèl·lules que van travessar el sistema fluídic per cada condició (*imatge 12*).



Imatge 12. Gràfics que representen la confluència cel·lular per diferents condicions de cada línia cel·lular. A) Línia cel·lular CML75 B) Línia cel·lular MLg. S’hi representen totes les condicions i cada una d’elles té tres gràfics que corresponen a diferents poblacions i una taula on apareix la confluència de cada població. Les cèl·lules tractades amb Cisplatí són les que tenen una major confluència, i les que tenen menor confluència són les cèl·lules control.

- qPCR

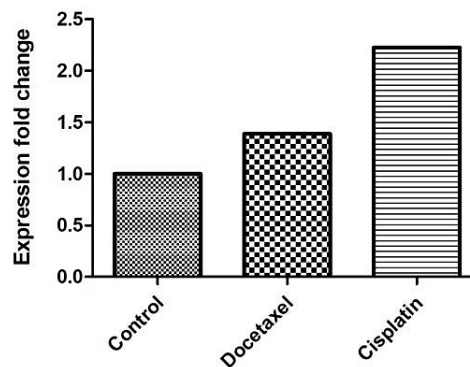
Les cèl·lules es van deixar sota tractament quimioterapèutic durant 7 dies. Passat aquest temps, es van rentar amb PBS, se’n va extreure l’ARN i es va convertir a cADN.

Es va utilitzar un plat de 96-pous per tal de poder-ho analitzar amb l’instrument CFX96 (BioRad). A cada pou es va addicionar SYBR green mix el qual conté els components necessaris perquè la reacció tingui lloc, els primers tant *forward* com *reverse*, aigua miliQ (*Ultrapure*), i la mostra de cADN. Els resultats obtinguts es mostren a la *imatge 13*.

| | | CTL | Cisplatin | Docetaxel | CTL | Cisplatin | Docetaxel | Water |
|-----|----|------------|------------|------------|------------|------------|------------|------------|
| | | A | B | C | D | E | F | G |
| p16 | 1 | 39.642 | 40.018 | 35.961 | ndetermine | ndetermine | 33.232 | ndetermine |
| | 2 | ndetermine | 36.757 | 36.011 | ndetermine | ndetermine | ndetermine | ndetermine |
| | 3 | ndetermine | ndetermine | ndetermine | ndetermine | ndetermine | ndetermine | 40.93 |
| | 4 | 35.81 | 35.93 | 35.777 | ndetermine | 39.102 | 34.904 | 33.541 |
| p21 | 5 | 34.729 | 35.862 | ndetermine | ndetermine | 36.769 | 34.807 | ndetermine |
| | 6 | ndetermine | 35.925 | ndetermine | ndetermine | 36.012 | ndetermine | ndetermine |
| | 7 | ndetermine | ndetermine | 28.845 | 34.637 | 23.943 | 25.628 | ndetermine |
| IL6 | 8 | ndetermine | ndetermine | 28.736 | 30.897 | 23.991 | 25.408 | 41.982 |
| | 9 | ndetermine | ndetermine | 28.852 | 30.282 | 23.708 | 25.301 | 34.359 |
| AKT | 10 | 34.501 | 34.501 | 22.551 | 28.274 | 16.015 | 16.632 | 34.222 |
| | 11 | 34.045 | 34.045 | 20 | 24.772 | 15.732 | 17.829 | 34.505 |
| | 12 | 34.927 | 34.927 | 19.638 | 25.688 | 19.35 | 23.372 | 34.735 |
| | | | | CML75 | | MLG | | No cDNA |

Imatge 13. Taula de valors obtinguts en l'assaig de qPCR. Cada fila correspon als gens que es volen analitzar en triplicats i cada columna correspon a la condició en la qual es trobaven les cèl·lules. Els gels a analitzar són biomarcadors de senescència i l'Actina (AKT), considerat un "House-keeping gene", es va utilitzar com a control. De la columna A fins a la C correspon a la línia cel·lular CML75, essent A el control, B el tractament amb Cisplatin i C el tractament amb Docetaxel. De la columna D a la F corresponen a la línia cel·lular MLG, essent D el control, E el tractament amb Cisplatin i F el tractament amb Docetaxel. La columna G correspon a un control complementat amb aigua. Tant els valors indeterminats com els superiors a 32 no es van considerar per l'anàlisi d'expressió, ja que podrien o no contenir mostra de cADN o ser nuls.

Només es va representar l'expressió de la Interleucina-6 (imatge 14), ja que és l'única que tenia uns valors extrapolables. Creiem que això va ser fruit dels primers.



Imatge 14. Expressió del gen que codifica per la interleucina-6 (IL-6) segons el seu fold-change per la línia cel·lular MLG i per cada condició aplicada. Tots els càlculs eren en log₂, de manera que cada vegada que hi ha el doble d'ADN, els valors de Ct disminueixen un 1 i no es reduïren a la meitat. Motiu pel qual s'han d'expressar els resultats d'expressió en fold-change a l'eix de les "y" com es mostra en el gràfic.

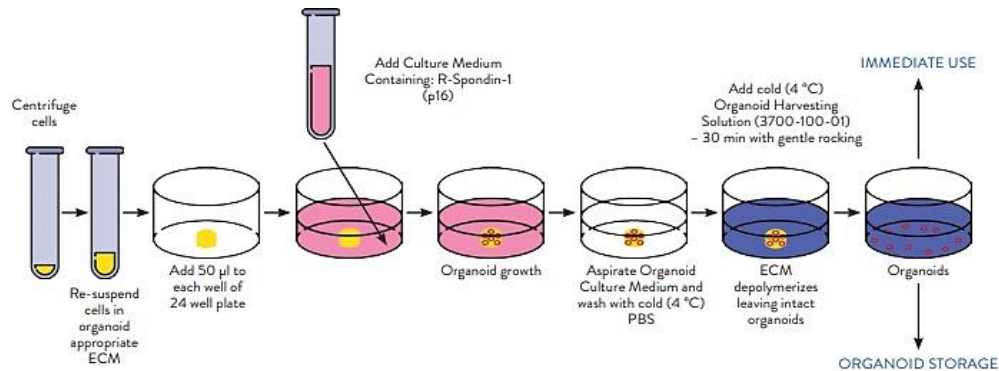
- Assaig d'Organoides

Aquesta tècnica in-vitro determina l'agregació de les cèl·lules en 3D permetent que s'autorenovin i s'organitzin en tres dimensions sense la funcionalitat d'un òrgan.

Amb aquesta metodologia, es volia analitzar si les cèl·lules senescents en condicions 3D tenien un efecte en mida o nombre d'organoides diferents segons la condició i la línia cel·lular.

En un plat de 24-pous, tal com es mostra a la *imatge 15*, es van sembrar cèl·lules immerses en un *matrigel*. Es va deixar solidificar la gota de matrigel que contenia les cèl·lules durant 20 minuts a l'incubador a 37°C.

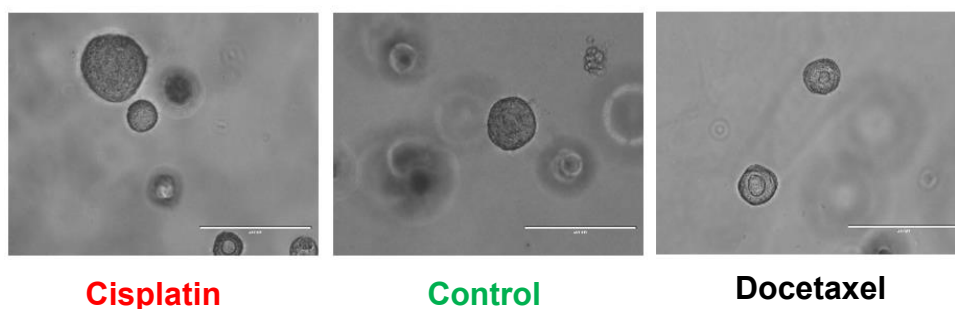
Un cop la gota aparentment es va solidificar, es van sembrar cèl·lules de cada condició al voltant de la gota omplint el pou i es van deixar créixer els organoides durant un període de set dies.



Imatge 15. Esquematzació del procés realitzat per l'anàlisi d'organoides Aquest va ser el protocol que es va seguir durant l'assaig d'organoides de cada una de les línies cel·lulars i per cada condició.

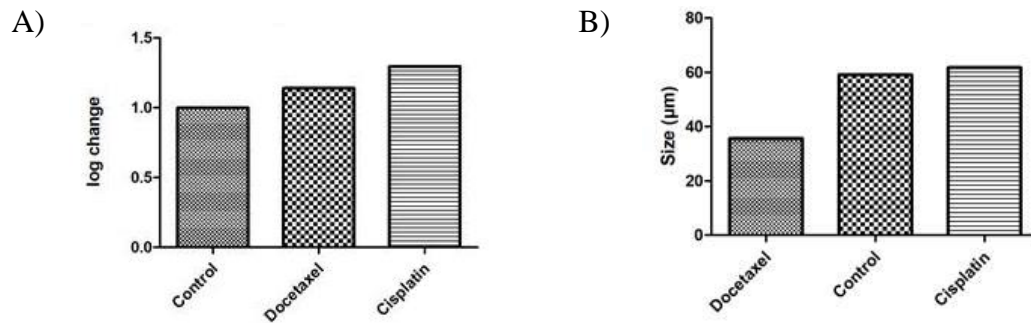
Amb aquest experiment es pretenia veure si les cèl·lules senescentes secreten de manera paracrina en el seu SASP, algun component que indueixi les cèl·lules normals proliferar de manera més descontrolada, autorenovar-se i organitzar esferoides 3D d'una manera més eficient.

Així doncs, es pretenia determinar si les cèl·lules induïdes a senescència per mitjà d'un component quimioterapèutic contribueixen en el creixement del tumor o la proliferació d'aquest. Els resultats es mostren a la *imatge 17*.



Imatge 16. Organoides de CML75 capturats per un microscopi convencional a 20X. Cada condició està esmentada i representada amb un color. Com s'observa, els organoides amb un major diàmetre són aquells que van ser tractats amb el medi col·leccionat de les cèl·lules induïdes a senescència per Cisplatí, i els de menor diàmetre són aquells que es van tractar amb el medi col·leccionat de les cèl·lules induïdes a senescència per Docetaxel.

Per mitjà del software del programa *ImageJ* es van quantificar els organoides totals per cada condició i es van representar a la *imatge 18a*). A més, també es va mesurar la mida dels organoides i es van representar per condició a la *imatge 18b*).



Imatge 17. Quantificació i mesura d'organoides de cèl·lules CML75. A) Quantificació en log-change dels organoides de CML75 per cada condició B) Mida en μm dels organoides de CML75 per cada condició. Com es pot observar en el primer gràfic, la presència de més organoides, són els que van ser tractats amb el medi col·leccionat de les cèl·lules induïdes a senescència per mitjà de Cisplatin. A més, si observem el segon gràfic, podem veure com els organoides de major diàmetre són aquells que van ser tractats amb el medi col·leccionat de cèl·lules induïdes a senescència per mitjà de Cisplatin, tot i que la diferència entre aquests i els de control no és molt representativa.

A.2. Evidències computacionals obtingudes durant el projecte d'anàlisi de dades de RNA-seq

| | logFC | logCPM | LR | PValue | FDR |
|-----------|-------------------|-------------------|------------------|----------------------|-----------------------|
| MSMP | -2.24759776539053 | 2.53427652842193 | 156.417288443507 | 6.86248783981373e-36 | 1.02244206325385e-31 |
| CUL4A | -1.06465604610093 | 6.81243144388149 | 151.412693764143 | 8.51506218229911e-35 | 6.34329557270373e-31 |
| PI3 | -2.96641444564115 | 4.10020084552877 | 138.067239722223 | 7.04504968394237e-32 | 3.49880650803525e-28 |
| CYP2B6 | -2.94108607878637 | 0.731454556957044 | 130.475100507965 | 3.22552996608553e-30 | 1.20142927411771e-26 |
| NPY | -6.79953637924709 | 2.43609353890901 | 109.937792979765 | 1.01114586439642e-25 | 2.51084370560705e-22 |
| CALB1 | -4.56291617020398 | 4.29265334691094 | 97.2191269629668 | 6.20642170625858e-23 | 1.32099252859352e-19 |
| MSTN | 4.14540510022863 | 1.45010436073434 | 93.8595229411857 | 3.36730727289179e-22 | 6.30843638235185e-19 |
| DPCR1 | 4.08611643239507 | 7.00925364334106 | 92.9515265639079 | 5.3590950084859e-22 | 8.87168405904793e-19 |
| LEFTY1 | -1.68517317103952 | 0.690096420204387 | 88.640891967978 | 4.73401304243856e-21 | 7.05320603192921e-18 |
| CACNA1A | -1.72422891922404 | 0.609805001215126 | 86.1500898686054 | 1.66784341763407e-20 | 2.25901809812091e-17 |
| CTCF | -1.43092798323116 | 0.396439014796355 | 80.9050738517274 | 2.36824826195233e-19 | 2.9403757123565e-16 |
| PAX8 | -1.35190058506319 | 2.5248061479498 | 75.4173312708195 | 3.81030715657002e-18 | 4.36690510197975e-15 |
| ALB | 5.98281111324178 | 5.97261076430484 | 69.9239032226404 | 6.16369902847712e-17 | 6.15412629561553e-14 |
| LOC283867 | 2.65367585821253 | 0.990977343818819 | 69.9136424206467 | 6.19584498518243e-17 | 6.15412629561553e-14 |
| DUOX2 | 2.19447838023063 | 4.34544479865157 | 60.8192811759817 | 6.25621597961603e-15 | 5.8257101175187e-12 |
| KRT13 | -2.28889843932713 | 3.41031878159801 | 59.0605218782498 | 1.5289619041647e-14 | 1.34000200509705e-11 |
| NLGN4X | -1.16420576005113 | 1.7439571918562 | 55.8816592545992 | 7.69672523838354e-14 | 5.73367546633382e-11 |
| LRRCS5 | 1.9092269078089 | 1.59478451851847 | 52.4369523748004 | 4.4428480058535e-13 | 2.87800083727875e-10 |
| TCHH | -1.28866280287277 | 0.952920073963788 | 51.2846985854953 | 7.98959687442259e-13 | 4.95087515966759e-10 |
| HIST1H4J | 1.90768278634945 | 2.86665483780127 | 51.1752681604572 | 8.44762574158446e-13 | 4.96214923993344e-10 |
| CUZD1 | -1.19868937002656 | 1.28223264737708 | 51.1266722912167 | 8.6593650740499e-13 | 4.96214923993344e-10 |
| ALDOB | 2.28969116268115 | 1.64478447476011 | 46.413107479934 | 9.57726979252618e-12 | 4.92040491858095e-09 |
| DUOXA2 | 1.76848492828998 | 2.02648613692112 | 45.9074231597058 | 1.2397529284201e-11 | 6.15702629351034e-09 |
| KRT96 | -1.15732727408924 | 1.92166085046035 | 45.2843394007788 | 1.70406313809745e-11 | 8.18994732081095e-09 |
| ZBTB20 | 1.54026595268717 | 1.09577129393768 | 44.80934457156 | 2.17183738960209e-11 | 1.01058907941916e-08 |
| FABP7 | 4.3425464719026 | 2.55324596391505 | 43.2647592616652 | 4.7811758135463e-11 | 2.0951393660596e-08 |
| CG5B | -3.18586481544171 | 1.14929170932152 | 42.9246807981098 | 5.68884126139685e-11 | 2.42165845581576e-08 |
| HIST4H4 | 1.52313129327655 | 1.39793119711732 | 42.6539470135772 | 6.5324661038386e-11 | 2.70385670133636e-08 |
| COL11A2 | 1.91683879990461 | 2.60328398449254 | 42.3812936693494 | 7.51049530250284e-11 | 3.02429377059432e-08 |
| HIST2H3C | 2.5135785122493 | 2.2931513600686 | 39.9731433148637 | 2.57478747578386e-10 | 9.83634835941122e-08 |
| H19 | -1.61012535624061 | 9.33478538388007 | 39.5846211298243 | 3.14147460804179e-10 | 1.17012075463037e-07 |
| CG6 | -2.63873037928277 | 1.16038805530667 | 38.9811631016942 | 4.27914617780503e-10 | 1.55499997324676e-07 |
| SCARNA2 | 2.79118013456829 | 4.84984294824733 | 37.8104550402229 | 7.79628260158027e-10 | 2.76563844002249e-07 |
| CDR1 | 2.331601843072 | 2.31868733054652 | 36.1685258447189 | 1.80969875812479e-09 | 6.27039576681425e-07 |
| SUCNR1 | 1.22331749446468 | 1.52420923976054 | 35.9127056335959 | 2.06358646812607e-09 | 6.98758517922962e-07 |
| HIST1H2BE | 1.25842255052452 | 0.781206513120829 | 34.393988972567 | 4.50111291047984e-09 | 1.457871331559216e-06 |
| A2ML1 | -2.57118946375015 | 2.2955164895787 | 34.2803697976132 | 4.77170636825932e-09 | 1.5126309187382e-06 |
| CLGN | -1.14512738798075 | 1.83456804628166 | 33.9603667270521 | 5.62481993465874e-09 | 1.7458589180168e-06 |
| PCSK2 | 2.72378856519841 | 6.00176750789027 | 33.5435159792407 | 6.96872976551286e-09 | 2.11892050564033e-06 |
| S100A7 | -2.89919515918872 | 1.9469990190306 | 32.6867672547163 | 1.08271825791845e-08 | 3.1630233970053e-06 |
| CEACAM4 | -1.64005210095096 | 3.73951991646619 | 32.5334169828974 | 1.17160567300573e-08 | 3.29353828719101e-06 |
| PRG4 | 1.59036148224303 | 5.11322452006333 | 32.3200015924681 | 1.30760389411902e-08 | 3.60777600342209e-06 |
| VTN | 1.86371180634407 | 3.4873954158175 | 31.9344287942209 | 1.5946554563743e-08 | 4.28323205031759e-06 |
| SCARNA9 | 1.51757665111853 | 0.709500669161887 | 31.4140939502342 | 2.08462188848215e-08 | 5.35496233043027e-06 |
| MUC2 | -2.35198883631217 | 1.71880444819557 | 30.5706656939327 | 3.21924475312187e-08 | 7.99392126279378e-06 |
| MAEL | 2.50916267061611 | 1.65628253696854 | 30.5065381999576 | 3.32743210367379e-08 | 8.04202410472783e-06 |
| IGSF9B | 1.24032448143116 | 1.85189610100908 | 30.4763133592193 | 3.37967901930387e-08 | 8.04202410472783e-06 |
| PKHD1L1 | 1.30714139220103 | 1.79898214412668 | 30.310079902416 | 3.68207440198755e-08 | 8.571754140010197e-06 |
| FGF12 | -1.03096231018688 | 1.34731954733825 | 28.6238605669685 | 8.7892454934929e-08 | 1.95449206676941e-05 |
| PAR5 | 1.10963499610943 | 0.972144758428267 | 28.2984321374378 | 1.03980170383799e-07 | 2.25589207793552e-05 |
| SCGB3A1 | 1.57743666618071 | 8.80430962196193 | 28.2892513628009 | 1.04474497199511e-07 | 2.25589207793552e-05 |
| SFRP1 | -1.05750595814519 | 3.32006332564858 | 28.1375984139206 | 1.12989356063276e-07 | 2.37102593800951e-05 |
| PNCK | -1.37622414726044 | 0.93709192328047 | 28.0735709500788 | 1.16789935549855e-07 | 2.41674062466291e-05 |
| LGR5 | -1.14522352248865 | 2.23440742513581 | 27.8596903496098 | 1.30623313806333e-07 | 2.66596815937337e-05 |
| SCARNA12 | 1.96081538486811 | 2.53491448327328 | 27.6963786492684 | 1.41928153002243e-07 | 2.85755074537895e-05 |
| ITLN1 | 2.54747001302235 | 5.14053744877722 | 26.8023850761759 | 2.25361359350898e-07 | 4.30469088842183e-05 |
| CYP2A6 | -1.81959745500903 | 2.40136388846419 | 26.5401467516958 | 2.58117850829143e-07 | 4.68987543841878e-05 |
| NKD1 | 1.03887850387324 | 2.49519958330166 | 26.4875237774839 | 2.65245667656611e-07 | 4.76131952098295e-05 |
| TDRD1 | -1.48505028302023 | 0.929103383831162 | 26.1941909274263 | 3.08749922047006e-07 | 5.47626796259326e-05 |
| STAC2 | 1.41783317851641 | 1.14117368032491 | 26.1015825292834 | 3.2391842136092e-07 | 5.67771830571334e-05 |
| MALAT1 | 1.12369954385333 | 9.58744935041021 | 26.0101451452882 | 3.39627946966457e-07 | 5.88385672308517e-05 |
| KRT4 | -1.27719634286983 | 3.67295083712525 | 25.9826299804618 | 3.44503066989109e-07 | 5.89971406356091e-05 |
| CCDC144A | 1.01718479989183 | 1.0497322393114 | 25.1486648700725 | 5.30763871554075e-07 | 8.5559713413084e-05 |
| HIST1H1B | 3.17967719904265 | 2.47569436200027 | 24.9392995248915 | 5.91639940080613e-07 | 9.2787825971169e-05 |
| MUC6 | 2.15062106706384 | 6.22536400247729 | 24.8810764178877 | 6.09780679800929e-07 | 9.46366911286879e-05 |
| PLA2G2A | 1.48950426774679 | 3.695341688493 | 24.7276471949335 | 6.60301569236675e-07 | 0.0001134769566188 |
| FAIM2 | 1.02741322747495 | 2.15668461529183 | 24.6157777090117 | 6.99764492664664e-07 | 0.000105860448602617 |
| SCARNA17 | 1.42928162913502 | 1.63355799802424 | 24.5159498022377 | 7.36972417734129e-07 | 0.000109801520518208 |
| GLDC | -1.04194519257076 | 2.43374261298896 | 24.4185471602379 | 7.75187902864358e-07 | 0.0001134769566188 |
| OLFM4 | -1.94300363902666 | 4.16415977807418 | 24.414360710532 | 7.76874258347376e-07 | 0.0001134769566188 |
| PGLYRP4 | -1.17360336085448 | 0.693051764132589 | 23.5479754207721 | 1.21837558986003e-06 | 0.00017250735031364 |
| DNAH11 | 1.07836994079913 | 4.35917312620973 | 23.5063273029571 | 1.24503639880869e-06 | 0.0001715738246158 |
| NPTX1 | 1.36187496915929 | 4.98664234408869 | 23.2806127543413 | 1.41467645600821e-06 | 0.000191611495628084 |
| UGT2B15 | 1.76780917444788 | 2.37736595552204 | 22.9627178080348 | 1.65173826749328e-06 | 0.000215870600415635 |
| FDPSL2 | 1.08077007341069 | 1.11821498933048 | 22.796907215699 | 1.80055230452361e-06 | 0.00023327383783454 |
| MYBP1C | 2.64951738635935 | 2.84468910630313 | 22.6545962776266 | 1.93995837493085e-06 | 0.000249039145059782 |
| DEF5A | 6.94423024141219 | 1.70438963764013 | 22.618575056957 | 1.97565256713678e-06 | 0.000251583312801461 |
| HIST1H2AM | 1.51194227317093 | 0.888600675495943 | 22.5982963391326 | 1.9966157828201e-06 | 0.000252098123290411 |
| STMN2 | 1.44547454182383 | 1.25834383980077 | 22.5553602614049 | 2.04174046146398e-06 | 0.000253919446038615 |
| LRRCS1 | 1.17046332997794 | 2.69355592526221 | 22.1298735132414 | 2.54813422526469e-06 | 0.000313757453076187 |
| B3GALT5 | 2.31376199159034 | 0.911638534618133 | 21.9895735883791 | 2.74135642994411e-06 | 0.000332060727233636 |
| SPINK5 | 1.13518770400393 | 4.81606281135569 | 21.6691601528722 | 3.23957880979082e-06 | 0.000386131877496587 |
| LOC728264 | 1.10866314817964 | 4.89160858943292 | 21.6483513136828 | 3.27491374265478e-06 | 0.000387245554379472 |
| WNT16 | -1.29440888023274 | 1.7191090461297 | 21.4785858663549 | 3.57802052405723e-06 | 0.000413247502232005 |

| | | | | | |
|-----------|--------------------|--------------------|------------------|-----------------------|----------------------|
| VGLL1 | -1.07186627179212 | 2.08667205544109 | 21.318178655773 | 3.89024852600722e-06 | 0.000442448952587646 |
| MIA | 1.60125741574968 | 1.52188960691913 | 21.0925625571372 | 4.37623367914405e-06 | 0.000484059318626945 |
| PRR4 | 1.13403537494839 | 2.83916125055187 | 20.8365743639666 | 5.0018607699697e-06 | 0.000540019736317236 |
| SCARNA7 | 1.95086737975985 | 2.14249342158902 | 20.6885007426526 | 5.4606980426876e-06 | 0.000581135292189359 |
| AFP | 5.92560260515315 | 2.01740749494427 | 20.4877188957958 | 6.00150923411106e-06 | 0.000629693563936765 |
| INHA | -1.33428368295916 | 3.70602933954245 | 20.4529847759788 | 6.11141723739202e-06 | 0.000632319482082665 |
| TKTL1 | 2.01991926368723 | 1.80857867390831 | 20.4029407288476 | 6.2733210607453e-06 | 0.000640180651014649 |
| WDR74 | 1.29739516190349 | 8.30256113927015 | 20.2001562303062 | 6.97472391382915e-06 | 0.000692776077280937 |
| TMSF4 | 1.74874114025043 | 5.24892960415814 | 20.179122260348 | 7.05183996666829e-06 | 0.000695797113053582 |
| PRB3 | 1.82153670366372 | 0.820959448580202 | 20.1084388230394 | 7.31731023090396e-06 | 0.000712789794530996 |
| NWD1 | 1.1097914636985 | 2.78451386204784 | 19.905437062949 | 8.1368586668488e-06 | 0.000762459479731952 |
| AMBP | 1.52898681985404 | 3.26461305184325 | 19.869878206536 | 8.28962314758005e-06 | 0.000764059009213722 |
| TF2 | 3.96018899050415 | 5.18680220103259 | 19.7765852300085 | 8.70424802736051e-06 | 0.000786566041203124 |
| SCN3A | -1.11043705990332 | 3.00439206936891 | 19.6364118154269 | 9.3667051818503e-06 | 0.000820990661790516 |
| TERC | 2.78273642008968 | 1.79633385936589 | 19.3246964969124 | 1.10270834476014e-05 | 0.000844209863726998 |
| LGALS7 | -1.23194090799046 | 0.779880409716991 | 19.2718360677447 | 1.13366331407314e-05 | 0.000965168555221467 |
| ORM2 | 1.17375788422911 | 2.80772830465497 | 19.2212827036996 | 1.16406331077147e-05 | 0.000979868771027356 |
| BMX | 1.07847752669866 | 1.29067684102917 | 17.9154149531645 | 2.30942985286406e-05 | 0.00147460890242749 |
| CLDN2 | 1.21569021539366 | 5.61591913856592 | 17.6084530265477 | 2.71379406316354e-05 | 0.0019900417347459 |
| HIST1H1E | 1.68248641694959 | 3.35947355996149 | 17.5065885172539 | 2.86313776726612e-05 | 0.00207077133953873 |
| HIST1H2AG | 1.64641976117883 | 1.55425159695723 | 17.3599472190296 | 3.09275281100774e-05 | 0.00219761788809963 |
| MMP10 | 1.28515498163055 | 3.25703136932482 | 17.2604106548586 | 3.25909609886246e-05 | 0.00223603989409611 |
| GRIN2A | 1.28473378264755 | 1.60392497691403 | 17.2289768416082 | 3.31343752441173e-05 | 0.00224395026800956 |
| KRT20 | 3.10876315848484 | 2.15296106866004 | 17.0146694493812 | 3.70921374970427e-05 | 0.00244529095826743 |
| PSCA | 1.43095363529423 | 5.01475050744406 | 16.8426926720165 | 4.06093849400129e-05 | 0.00260792769922953 |
| MUC17 | 5.59357005897301 | 1.5530041196441 | 16.8160380926373 | 4.11837677050613e-05 | 0.0026220290956268 |
| DKK4 | 5.51263881247117 | 0.611187172745334 | 16.6683934136512 | 4.45165495531424e-05 | 0.00279853194849058 |
| LECT1 | 2.90301887949737 | 4.21380855468416 | 16.5856945174519 | 4.65004529916057e-05 | 0.0028747313241574 |
| CPB1 | 2.31671614277343 | 1.0750556752998 | 16.4266110357422 | 5.05702833144011e-05 | 0.00305224851847654 |
| DAZ1 | 5.45400088563807 | 1.5612975629196 | 16.231359422568 | 5.60584402481678e-05 | 0.00326255742678692 |
| PCDH8 | 2.76838084639485 | 1.09665381212386 | 16.1596664458797 | 5.822068456742e-05 | 0.00336213170298446 |
| CYP11A1 | 2.5285995039479 | 2.40395587959483 | 16.0105564943325 | 6.29902776814208e-05 | 0.00356841120599045 |
| WNA3A | 1.05731471499893 | 1.63896421579961 | 15.418161631537 | 8.61562640948779e-05 | 0.00428663398163808 |
| HOXB8 | -1.100448162241571 | 1.42108182133598 | 15.3505518765879 | 8.92948409140873e-05 | 0.00434772945025812 |
| CST4 | 1.44746951390134 | 0.596710598973341 | 15.2838024127868 | 9.2506347145779e-05 | 0.00442172932760389 |
| WFDC12 | 1.44420512776784 | -0.152411872684003 | 15.0565756993884 | 0.000104336126491466 | 0.0048883719684388 |
| FER1L6 | 1.97650552951775 | 2.81427938311674 | 15.0394083062694 | 0.000105289464097967 | 0.00490221164248631 |
| SYT4 | 2.00549530498046 | 1.82644894212509 | 14.712511896144 | 0.000125212653660582 | 0.00552138762468153 |
| CYP17A1 | 1.65590850721427 | 1.15534426721701 | 14.4045160916519 | 0.000147448274658933 | 0.00520102906882234 |
| SCGB3A2 | 1.01210155638216 | 8.70925435314808 | 14.2718702823265 | 0.000158211666561705 | 0.00651159011078132 |
| AP0H | 1.08169806053473 | 3.57801285414094 | 13.8897498982917 | 0.000193852801060229 | 0.00734206045709148 |
| CALML3 | -1.63080001399879 | 3.21255512351639 | 13.5734580248223 | 0.00022940636354208 | 0.0095054876419857 |
| SNORA57 | 2.75368413028725 | 3.75436475233522 | 13.2337427507177 | 0.000274954237602854 | 0.00918946734100797 |
| ERN2 | 1.07666782812716 | 4.48568987944886 | 13.2016658883391 | 0.00027970038634564 | 0.0092722467880137 |
| MYO1A | 1.02367578512153 | 0.661790687188437 | 12.5526963051938 | 0.000395635029320252 | 0.011626365144821 |
| HIST1H2AH | 2.64658561680463 | 0.735693721319064 | 12.5442495893001 | 0.000397427436253073 | 0.0116401136204593 |
| SNORD97 | 2.61266390316001 | 2.05700712131778 | 12.454535789448 | 0.0004147697994403696 | 0.0119472214202444 |
| TTR | 1.8811875660789 | 1.38833310423829 | 12.2700681318986 | 0.000460281466134192 | 0.0128027383487877 |
| SPINK1 | 1.13778782810842 | 7.62745513097954 | 12.0595104309 | 0.000515288224693316 | 0.0136849897677464 |
| NTS | -1.35852549514446 | 5.82752958546204 | 11.8388776550553 | 0.000580067777500164 | 0.0147230409919505 |
| FGA | -1.06714597583842 | 8.82946109195216 | 11.8567026675584 | 0.000639716267141031 | 0.0155290495132651 |
| A0H6 | 1.04080120703882 | 0.862147637677923 | 11.6131464679198 | 0.00065487269092348 | 0.0157116718551835 |
| PGC | 1.18879741424005 | 9.9676907429304 | 11.5550005107474 | 0.00067567289606957 | 0.016006954584544 |
| SCGB2A1 | 1.03333615492315 | 1.67221186549097 | 11.460104665521 | 0.000711062950162869 | 0.016425002937173 |
| SNORD15B | 2.5877247947239 | 1.2711435326525 | 11.2924146208323 | 0.000778243989782707 | 0.0175416901721219 |
| FLG1 | -1.07663897158769 | 6.048292206135 | 11.2441176858781 | 0.000798757541325854 | 0.0177887722095873 |
| ILF3 | -1.01873030241846 | 1.88190948181809 | 10.776490040418 | 0.00102797429114907 | 0.0210961280439526 |
| ANXA10 | 2.03408041464283 | 3.31338232983676 | 10.71336708060 | 0.00106364287556644 | 0.021498481448745 |
| RMRP | 1.44835551180508 | 8.87092421232447 | 10.671436095213 | 0.00108802613198749 | 0.0215565177400022 |
| TFAP2B | 1.82205236729686 | 0.641218554597327 | 10.6381182301312 | 0.00110780337176256 | 0.0217815405028959 |
| D5G3 | -1.38375636643425 | 1.91264350601099 | 10.5604717709709 | 0.00115531705687998 | 0.022412850395648 |
| CP51 | -1.04156091925127 | 8.21612702556809 | 10.4127898456271 | 0.00125145540256697 | 0.0238313751638293 |
| HIST1H1D | 1.4308769179411 | 2.24802869555096 | 10.2071914777429 | 0.00139894054280024 | 0.0253254133015562 |
| RPPH1 | 1.36065528252895 | 8.95898388651458 | 9.95512804067812 | 0.00160401970403467 | 0.0275009085956523 |
| SCARNA6 | 1.82456398288193 | 3.00627526360125 | 9.88734856475935 | 0.00166418998824355 | 0.0282400515876203 |
| C20orf70 | -1.0521775634159 | 1.97805817449812 | 9.08184495251987 | 0.00205815947390331 | 0.0363890066384610 |
| KRT14 | 1.2745975273039 | 4.03805336192401 | 9.00165954568035 | 0.00269734557476904 | 0.0374066833876875 |
| TOU3F2 | -1.27623244941747 | 2.57895087132367 | 8.64931068609337 | 0.00327185116465889 | 0.0417742963840957 |
| PRIM48 | -2.32140985207885 | 0.559213306129512 | 8.53702187254029 | 0.00347994548874291 | 0.043387203210695 |
| NTSR1 | -1.00824591564901 | 1.5377877733417 | 8.46087548923686 | 0.0036286715164226 | 0.04442364578373298 |
| VCX3A | -1.10918596340838 | 0.549232344066217 | 8.33343761078777 | 0.00389219370481549 | 0.0463177268434873 |
| SCGB1A1 | 1.00876828628031 | 8.38760804187891 | 8.31558155399341 | 0.00393064171415346 | 0.0464401433551528 |
| SCGB1A1 | 1.00876828628031 | 8.38760804187891 | 8.31558155399341 | 0.00393064171415346 | 0.0464401433551528 |
| ANXA13 | 1.28204602860808 | 0.938745080695941 | 8.19684745363816 | 0.00419632417091358 | 0.0481744898061959 |
| SPRR1B | -1.03718285116073 | 2.85899715721446 | 8.162556340577 | 0.00427640527824373 | 0.0486471247455049 |
| SNORD10 | 1.74154148081956 | 3.1579016415661 | 7.99512072261336 | 0.00469035728935384 | 0.0510120547348881 |
| KCNQ2 | 1.66218275426483 | 0.566149374099568 | 7.90272340269689 | 0.00493604249089285 | 0.0526114270899075 |
| GC | 3.34401643097306 | 1.57925757910002 | 7.89357881162337 | 0.00496106036523866 | 0.0526114270899075 |
| HIST1H1D | 2.06283871849396 | 1.01330058586829 | 7.86570669911663 | 0.00503811319280234 | 0.0531230350032287 |
| HIST1H3B | 1.81429277769417 | 1.14232527800935 | 7.86010972573149 | 0.0050537325466976 | 0.0532051166946774 |
| PRSS1 | 1.33582519233569 | 2.69946971187274 | 7.78073561989629 | 0.00528062903823694 | 0.0545603969950951 |
| TME229A | 1.28319086155714 | 0.734890881718948 | 7.6162919457822 | 0.00578433032776652 | 0.055305324121451 |
| CDHR5 | 1.00685522056971 | 2.07064617065479 | 6.9432538580736 | 0.00841359258064508 | 0.0718591113408983 |
| SLC10A2 | 1.56489408139573 | 1.26881976804853 | 6.82891839587575 | 0.00896935636612085 | 0.0743652979960125 |
| SLC14A2 | 1.03873685081282 | 2.52977874287203 | 6.54527738825175 | 0.0105162930493442 | 0.0820325916974764 |
| SNORA71A | 2.92706540121405 | 0.874542582923517 | 6.41359888141633 | 0.0113249650867229 | 0.0853036677588901 |
| SCARNA10 | 1.28277133554537 | 4.65394739362708 | 6.15331762525318 | 0.0131165823190276 | 0.0933256733386783 |
| TMSF10 | -1.33522889701448 | 1.60735038648977 | 6.01484234165105 | 0.0141860458502794 | 0.0973102657105491 |
| SNORA81 | 1.6689536859936 | 0.976799747382224 | 5.87094884416774 | 0.0153927531681352 | 0.101207691726411 |
| CALCA | -1.17468813287993 | 7.76890375332049 | 5.80899505207117 | 0.0159444031666448 | 0.103061025067176 |
| KLK5 | 1.20980195815016 | 2.00335370752078 | 5.75121317421747 | 0.0164772739549461 | 0.104988576599453 |
| SCARNA5 | 1.43661428703074 | 4.40876785432969 | 5.70935273410794 | 0.0168747602022689 | 0.106487527426347 |
| SNORA63 | 1.62289775613196 | 4.32120457794906 | 5.5871112809258 | 0.018093110577159 | 0.109938521406644 |

| | | | | | |
|----------|-------------------|-------------------|------------------|--------------------|-------------------|
| SNORD17 | 1.11380455781237 | 4.09742611189727 | 5.54142611096847 | 0.0185714425589767 | 0.111525966419264 |
| CT45A1 | -1.41348539065701 | 1.23271396948118 | 5.49364446047997 | 0.0190857184666082 | 0.112840523584919 |
| INSL4 | -1.39869059751894 | 2.13548783238559 | 5.03337284625889 | 0.0248634328880686 | 0.1314778683265 |
| SNORA49 | 2.31532532867381 | 0.953366831219538 | 4.36469227683602 | 0.0366911095009946 | 0.162841860680691 |
| GAGE12D | -1.76597168595459 | 2.5238861995733 | 4.19571606713436 | 0.0405262346058618 | 0.17204944964946 |
| F2 | -1.33372263821759 | 0.669343420011452 | 4.02986862221314 | 0.0447014218365423 | 0.182818139978766 |
| CALML5 | -1.09062779784514 | 2.99041614882071 | 3.94587666612421 | 0.0469863656218886 | 0.187731258085417 |
| HIST1H4C | 1.14764246687838 | 1.50271752889125 | 3.79057483615128 | 0.0515419446226106 | 0.198892368021827 |
| SPINK4 | -1.20080997351076 | 3.56452506601638 | 3.73073160903198 | 0.0534199905533001 | 0.202674927235452 |
| REG1A | 1.42056662195301 | 2.67388902713923 | 3.68061667141734 | 0.0550484865945492 | 0.206728591842335 |
| REG4 | 1.09744303258897 | 4.23300641723848 | 3.61081132034951 | 0.0574051082322333 | 0.212755897401006 |
| SNORA74A | 1.9536934196276 | 1.08784407720662 | 3.2685502859124 | 0.070619800498447 | 0.241053681333963 |
| SNORA68 | 1.93068885760914 | 1.93130053650192 | 3.19676573534446 | 0.0737840509678838 | 0.248022588428749 |
| SNORA54 | 1.86488953590796 | 1.53821558661606 | 3.01264568113032 | 0.0826173424288976 | 0.265340759829305 |
| SNORA48 | 1.85216756606732 | 3.42565672434331 | 2.97092519608628 | 0.0847733586135769 | 0.269476908466755 |
| SNORA23 | 1.73307349971917 | 2.01787638744464 | 2.64892290863312 | 0.103620083805743 | 0.302594203963498 |
| MAGEA10 | -1.22870865839708 | 1.18379721256583 | 2.60875098967165 | 0.106275436123373 | 0.305972635411498 |
| SNORA74B | 1.70210481949836 | 1.3293140952749 | 2.56946873249416 | 0.108944313962938 | 0.310237257976647 |
| GCG | 1.6621642307484 | 0.933641729364985 | 2.46670327385789 | 0.116281589932864 | 0.32196235056862 |
| SNORA52 | 1.64869919314967 | 1.06673934003405 | 2.43064304401446 | 0.118984080310014 | 0.326773053002561 |
| SNORA70 | 1.4768109229071 | 1.80687456236975 | 1.99551333744486 | 0.15776560333313 | 0.385032177028585 |
| GAGE2B | -1.12654719659408 | 0.507342035160608 | 1.62108935550495 | 0.202939959390089 | 0.443538573412488 |

Imatge 18. Taula del resultat de l'anàlisi d'expressió diferencial del projecte TCGA-LUAD. Les files representen els gens i les columnes representen els seus valors pertinents de 'log fold-change (logFC)', log counts-per-milion (logCPM), log fold-change ratio (LR), p-valor (p-value) i p-valor ajustat (FDR). Els valors negatius presents a la columna de logFC és pel fet que, malgrat que els gens es filtressin per un valor major que 1, no vol discernir en signes, ja que els negatius representen els gens infraexpressats, mentre que els positius impliquen els gens sobreexpressats. En aquesta taula apareixen els 198 gens diferencialment expressats.