Implicación de las mutaciones del gen K-RAS en el cáncer de pulmón

Pere Pavón Vizcaino
Máster Universitario de Bioinformática y Bioestadística
Biología del desarrollo, cáncer, biología molecular y farmacología

Ivette Olivares Castiñeira
Carles Ventura Royo

Junio 2018

Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada 3.0 España de Creative Commons

FICHA DEL TRABAJO FINAL

Título del trabajo:	*Implicación de las mutaciones del gen K-RAS en el cáncer de pulmón*
Nombre del autor:	*Pere Pavón Vizcaíno*
Nombre del consultor/a:	*Ivette Olivares Castiñeira*
Nombre del PRA:	*Carles Ventura Royo*
Fecha de entrega (mm/aaaa):	06/2018
Titulación:	*Máster Universitario de Bioinformática y Bioestadística*
Área del Trabajo Final:	*Biología del desarrollo, cáncer, biología molecular y farmacología*
Idioma del trabajo:	*Castellano*
Palabras clave	*MAPK, KRAS, LUNG CANCER*

Resumen del Trabajo (máximo 250 palabras): *Con la finalidad, contexto de aplicación, metodología, resultados i conclusiones del trabajo.*

El cáncer de pulmón es uno de los cinco cánceres que producen más muertes en la población en Europa en la actualidad, cerca del 20,1%. Las mutaciones génicas en el gen RAS (K-RAS, N-RAS y H-RAS) afectan a los sistemas de señalización celular, dotando a las células tumorales de algunas características que hacen que su comportamiento se transforme en maligno. La vía MAPK/ERK es una de las vías de señalización, junto con la isoforma K-RAS mutada, que está implicada en la oncogénesis y desarrollo del Cáncer de Pulmón Célula No Pequeña (NSCLC).

Dada la complejidad del funcionamiento de la red de las vías de señalización celulares y sus líneas de retroalimentación, es necesario identificar proteínas/genes que tengan influencia en este tipo de cáncer (NSCLC) y en el funcionamiento de las terapias aplicadas a los pacientes que son portadores del gen mutado KRAS. Por ello, este trabajo se ha centrado en el estudio de la vía MAPK/ERK, en las mutaciones de KRAS y en la expresión de ocho genes implicados en estas vías, efectuando un estudio sobre 30 pacientes con este tipo de cáncer y KRAS mutado, analizando la influencia de cada gen en el pronóstico de los pacientes a través de un análisis exhaustivo de supervivencia. En el estudio se demuestra que las variaciones de las expresiones en algunos de los genes presentan evidencia significativa al 5% de que influyen en el tiempo de supervivencia de los pacientes que tienen este tipo de cáncer y en la efectividad de las terapias utilizadas.

Abstract (in English, 250 words or less):

Lung cancer is one of the five cancers that cause the most deaths in the population in Europe today, about 20.1%. Gene mutations in the RAS gene (K-RAS, N-RAS and H-RAS) affect cellular signaling systems, giving the tumor cells some characteristics that cause their behavior to become malignant. The MAPK / ERK pathway is one of the signaling pathways, together with the mutated K-RAS isoform, that is involved in the oncogenesis and development of Non-Small Cell Lung Cancer (NSCLC).

Given the complexity of the functioning of the network of cellular signaling pathways and their feedback lines, it is necessary to identify proteins / genes that have an influence on this type of cancer (NSCLC) and on the functioning of the therapies applied to patients who are carriers of the mutated KRAS gene. Therefore, this work has focused on the study of the MAPK / ERK pathway, on KRAS mutations and on the expression of eight genes involved in these pathways, carrying out a study on 30 patients with this type of cancer and mutated KRAS, analyzing the influence of each gene on the prognosis of patients through a comprehensive analysis of survival. The study demonstrates that the variations of the expressions in some of the genes present significant evidence at 5% that influence the survival time of patients with this type of cancer and the effectiveness of the therapies used.

Índice

1. INTRODUCCIÓN 1

1.1 CONTEXTO Y JUSTIFICACIÓN DEL TRABAJO 1
1.1.1 Descripción general 1
1.1.2 Justificación del TFG 2
1.2 OBJETIVOS DEL TRABAJO 3
1.2.1 Objetivos generales 3
1.2.2 Objetivos específicos 4
1.3 ENFOQUE Y MÉTODO SEGUIDO 4
1.4 PLANIFICACIÓN DEL TRABAJO 7
1.4.1 Tareas 7
1.4.2 Calendario 8
1.4.3 Hitos 9
1.4.4 Análisis de riesgos 12
1.5 RESULTADOS ESPERADOS 13
1.5.1 Plan de trabajo 13
1.5.2 Memoria 13
1.5.3 Producto 14
1.5.4 Presentación virtual 14
1.5.5 Autoevaluación del proyecto 14
1.6 ESTRUCTURACIÓN DE MEMORIA DEL PROYECTO 14

2. VÍAS DE SEÑALIZACIÓN CELULAR. 17

2.1 VÍAS MAP KINASAS 19
2.1.1 Vías de señalización MAPK 19
2.2 FUNCIONAMIENTO DE LA VÍA MAPK/ERK: FAMILIA RAS Y SU IMPLICACIÓN ONCOGÉNICA 22
2.2.1 Vía o cascada RAS-ERK 22
2.2.2 Familia RAS y su implicación oncogénica 25
2.2.3 Estructuras del dominio de las proteínas RAS 26
2.3 MUTACIONES EN KRAS: SUBTIPOS SEGÚN MUTACIÓN 27
2.3.1 Oncogenes RAS humano: HRAS, NRAS y KRAS. 27
2.3.2 Mutaciones en RAS. 29
2.4 DIANAS TERAPÉUTICAS RELACIONADAS CON KRAS 35
2.4.1 Diana: vías de activación de RAS. 36
2.4.2 Diana: vías efectoras de RAS aguas abajo. 37
2.4.3 Diana: interacción letal sintética. 38
2.4.4 Diana: Genes relacionados con vías de señalización y KRAS 38

3. ESTUDIO POBLACIONAL DE INCIDENCIA DEL CÁNCER DE PULMÓN 40

3.1 ESTADÍSTICA GLOBAL DE INCIDENCIA DE CÁNCER EN EL MUNDO Y EN EUROPA 40
3.1.1 Incidencia del cáncer en la por Sexo 41
3.1.2 Principales cánceres en la población europea 42
3.1.3 Incidencia y mortalidad del cáncer de pulmón por regiones OMS (WHO) 43
3.2 INCIDENCIA EN LA POBLACIÓN DE CÁNCERES RELACIONADOS CON MAPK/ERK 44
3.2.1 Análisis global de frecuencia de mutaciones génicas en la vía MAPK/ERK 44
3.3 COMPARATIVA DE LA INCIDENCIA DEL CÁNCER DE PULMÓN CON OTROS TIPOS DE CÁNCER 46
3.3.1 Ranking comparativo de frecuencias de tipos de cáncer con mutaciones génicas en la vía MAPK/ERK 46
3.3.2 Análisis comparativo de implicación de mutaciones de genes según el tipo de cáncer 47

4. ESTUDIO BIOESTADÍSTICO COMPARATIVO: 50

4.1 DESCRIPCIÓN DEL ESTUDIO ORIGINAL 51
4.2 ANALISIS BIOINFORMÁTICO Y BIOESTADÍSTICO 52
4.2.1 Objetivos del proceso de análisis *52*
4.2.2 Pipeline para el proceso de análisis *53*
4.2.3 Proceso de análisis *53*
4.2.4 Conclusiones del análisis de supervivencia *76*
4.2.5 Comparativa de resultados con estudio original *80*
5. CONCLUSIONES **83**
6. GLOSARIO **85**
7. BIBLIOGRAFÍA **88**
8. ANEXOS **92**
8.1 ANEXO I. INFORME DE ESTUDIO POBLACIONAL DEL CÁNCER DE PULMÓN 92
8.2 ANEXO II. INFORME DE ANÁLISIS DE SUPERVIVENCIA DE CÁNCER NSCLC. 92

Lista de figuras

Figura F 1.1. *Estrategia de análisis "Top-Down", parte investigación teórica.* 5
Figura F 1.2*. Planificación de tareas del TFM.* 10
Figura F 1.3. *Diagrama de Gantt del TFM.* 11
Figura F 2.1. *Proteínas en vía de señalización.* 18
Figura F 2.2. *(KEGG, Consultado en 2018) Vías de señalización celular MAPK* 21
Figura F 2.3. *(Figura adaptada de Dirk Bokemeyer, 1996) Esquema componentes de la ruta RAS-ERK* 23
Figura F 2.4. *(Figura 2A de Baines, 2011) Ciclo de activación de Ras normal* 24
Figura F 2.5. *(Figura 2B de Baines, 2011) Activación crónica de Ras mutado* 26
Figura F 2.6. *(Figura 1 de Hancock, 2003) Estructura de dominio de RAS* 27
Figura F 2.7. *(HGNC, 2018) Mapa de jerárquico de la familia de genes RAS* 28
Figura F 2.8. *(UCSC, 2018) Gen HRAS en el navegador genómico de UCSC* 28
Figura F 2.9. *(UCSC, 2018) Gen KRAS en el navegador genómico de UCSC* 29
Figura F 2.10. *(UCSC, 2018) Gen NRAS en el navegador genómico de UCSC* 29
Figura F 2.11. *(COSMIC v84, 2018) Distribución de mutaciones por sustitución en codones 12, 13 y 61.* 30
Figura F 2.12. *(COSMIC v84) Distribución de tipos de sustituciones en residuos G12 y G13.* 31
Figura F 2.13. *(Grafico adaptado de COSMIC v84) Sustituciones en G12 y G13 de KRAS.* 32
Figura F 2.14. *Mutaciones KRAS en Adenocarcinoma de pulmón y en NSCLC* 35
Figura F 3.1. *Regiones de la OMS (WHO)* 43

Lista de Tablas

Tabla 2.1*. Mutaciones somáticas de KRAS en codón 12 (A) y 13 (B) en cáncer.* 33
Tabla 3.1. *Distribución de Cáncer en el mundo según sexo* 41
Tabla 3.2 *Distribución de frecuencias de genes mutados* 45
Tabla 3.3. *Distribución de frecuencias de tipos de cáncer y de genes mutados* 47
Tabla 4.1. *Test de Shapiro-Wilk de expresiones génicas* 57
Tabla 4.2. *Datos de la muestra para el Análisis de Supervivencia* 60
Tabla 4.3. *Resultados del modelo de regresión de Cox Univariable.* 63
Tabla 4.4. *Modelo de Cox Univariable Estratificado (Alta, Media y Baja)* 64
Tabla 4.5. *Modelo de Cox Univariable Estratificado (Alta, Baja-Media)* 65
Tabla 4.6. *Modelo de Cox Univariable Estratificado <Q25 y >=Q25* 65
Tabla 4.7*. Estratificación: Mejores puntos de corte* 66
Tabla 4.8. *Modelo de Cox Univariable estratificado con mejores puntos de corte* 66
Tabla 4.9. *Variables significativas de cada modelo de regresión de Cox* 69
Tabla 4.10. *Significación global de cada modelo Multivariable* 69
Tabla 4.11. *Resultados del modelo de regresión de Cox Univariable (GL)* 72
Tabla 4.12. *Modelo de Cox Univariable Estratificado (Alta, Media y Baja) SGL* 72
Tabla 4.13. *Modelo de Cox Univariable Estratificado (Alta, Baja-Media) SGL* 73
Tabla 4.14. *Modelo de Cox Univariable Estratificado <Q25 y >=Q25 SGL* 73
Tabla 4.15 *Estratificación a partir de mejores puntós de corte* 73
Tabla 4.16*. Modelo de Cox Univariable estratificado con mejores puntos de corte* 74
Tabla 4.17. *Variables significativas de cada modelo de regresión de Cox SGL* 75
Tabla 4.18. *Significación global de cada modelo* 76

Lista de Gráficas

Gráfica Gr. 3.1 *Gráfica de datos de cáncer en el mundo.* 41
Gráfica Gr. 3.2. *Gráfica de datos de cáncer en Europa.* 42
Gráfica Gr. 3.3. *Gráfica de ranking de principales cánceres en Europa* 43
Gráfica Gr. 3.4. *Gráfica de Incidencia vs Mortalidad del cáncer de pulmón en mundo* 44
Gráfica Gr. 3.5*. Gráfica de distribu-ción de genes mutados.* 45
Gráfica Gr. 3.6. *Distribución de frecuencias de tipos de cáncer* 46
Gráfica Gr. 3.7*. Distribución de frecuencias de mutación de genes* 48

Gráfica Gr. 3.8. *Distribución de frecuencias de mutación de genes por tipo de cáncer* 49
Gráfica Gr. 4.1. *Distribución de las edades de los pacientes por género.* 54
Gráfica Gr. 4.2 *Hábio tabáquico* 54
Gráfica Gr. 4.3 *Consumo de tabaco* 54
Gráfica Gr. 4.4 *Distribución según histología del cáncer.* 55
Gráfica Gr. 4.5 *Distribución según mutación KRAS* 55
Gráfica Gr. 4.6. *Nivel de expresión génica* 56
Gráfica Gr. 4.7.*Test de correlación de Spearman.* 58
Gráfica Gr. 4.8 *Curva de Supervivencia SLP* 61
Gráfica Gr. 4.9. *Riesgo acumulado SLP* 61
Gráfica Gr. 4.10. *Curva de Supervivencia LKB1 estratificada* 64
Gráfica Gr. 4.11. *Curva de supervivència HES1 Estratificada* 65
Gráfica Gr. 4.12. *Curva de supervivència LKB1 Estratificada* 65
Gráfica Gr. 4.13. *Curvas de Supervivencia* 67
Gráfica Gr. 4.14. *Curva de Supervivencia GL* 71
Gráfica Gr. 4.15. *Riesgo acumulado SGL* 71
Gráfica Gr. 4.16. *Curva de Supervivencia HES1 estratificado* 74
Gráfica Gr. 4.17. *Curva de Supervivencia CDCP1 estratificado* 74
Gráfica Gr. 4.18. *Curva de Supervivencia LKB1 estratificado* 75

Capítulo 1

1. Introducción

1.1 Contexto y justificación del Trabajo

1.1.1 Descripción general

El presente trabajo se centra en el estudio del cáncer de pulmón cuya carcinogénesis está relacionada con mutaciones génicas en las vías de señalización celular MAPK/ERK y concretamente en el gen K-RAS.

Por una parte, se pretende describir y analizar el funcionamiento de las vías de señalización MAPK desde un enfoque más global, pasando posteriormente a un análisis más específico del funcionamiento de la vía MAPK/ERK y de las proteínas RAS, para llegar a entender como la mutación del gen K-RAS actúa como promotor del cáncer y en concreto del cáncer pulmón. Por otra parte, se analizará la mutación concreta que se produce en el gen K-RAS y que induce al funcionamiento no regulado de la actividad celular en estas vías.

Siguiendo la línea del gen diana K-RAS, se abrirá un ámbito de búsqueda de estudios realizados sobre análisis de dianas terapéuticas para este tipo de cáncer, para ver qué tipos de fármacos se han utilizado en ellos, así como los resultados obtenidos y efectos adversos que se hayan podido dar.

Adicionalmente, partiendo del gen diana estudiado, el K-RAS, se buscarán estudios de investigación biológica y/o de estadística publicados sobre la incidencia del cáncer en la población relacionados con la vía MAPK/ERK, para hacer una comparativa entre el cáncer de pulmón y otros tipos de cáncer relacionados con mutaciones genéticas en esta vía. De esta manera se mostrará la incidencia de este tipo de cáncer en la población versus otros tipos.

Por último, en base a la disponibilidad de datos de un estudio clínico actual, se efectuará un análisis bioestadístico de los datos recogidos para el estudio, sobre la incidencia en la vía de señalización analizada y con la mutación en el gen K-RAS activa en pacientes con cáncer de pulmón, comparando los resultados obtenidos con los obtenidos en ensayo original en alguno de los ámbitos analizados en el estudio original.

1.1.2 Justificación del TFG

El hecho de realizar el presente trabajo en el área de investigación del cáncer de pulmón, enfocado en el estudio de su oncogénesis sobrevenida por mutaciones génicas, concretamente en el gen K-RAS, está fundamentado básicamente en tres factores determinantes: motivación personal, relevancia de la temática en el ámbito de la investigación científica del cáncer y por último, el acceso a datos actuales de ensayos o estudios clínicos. Estos tres factores justifican tanto el por qué de su desarrollo, como sobre la temática elegida.

En primer lugar, cabe destacar la motivación caracterizada por el interés personal sobre la materia tratada en el trabajo, así como el reto que representa sumergirse en un proyecto de investigación biológica, abandonando o como menos, aparcando momentáneamente el área puramente tecnológico habitual del trabajo profesional. Ambos aspectos han sido un factor determinante para orientar el trabajo en esta dirección, que junto con la posibilidad de aprovechar la experiencia en abordar proyectos tecnológicos en el ámbito de la Salud, seguro aportará un valor añadido a resultado final del trabajo.

En segundo lugar, como otro factor de justificación y posiblemente más importante, sobresale el planteamiento de un trabajo de investigación sobre una temática que tiene una relevancia destacada actualmente en la comunidad científica de investigación sobre el cáncer, por la alta incidencia en la población que tiene esta enfermedad. En Europa, el cáncer de pulmón representa el 12,1 % (449 K) de nuevos casos de cáncer y el 20,1 % (388 K) de muertes relacionadas con el cáncer (IARC - World Health Organitation, 2012; NIH - National Cáncer Institute, 2018). Podemos observar que la proporción de muertes por cáncer de pulmón versus a los nuevos casos diagnosticados, es muy alta. En Europa representa el 86,4 % aproximadamente. Estos porcentajes de incidencia y mortalidad de este tipo de cáncer hacen que una parte importante de los proyectos de investigación científica sobre el estudio del cáncer en la actualidad se enfoquen en el estudio de sus causas, identificación de dianas terapéuticas efectivas y tratamientos para reducir o eliminar sus efectos.

Para evidenciar el interés creciente en los últimos años de estudios realizados sobre el cáncer de pulmón por la comunidad científica, se ha realizado una búsqueda en PubMed (desde el 2010 hasta la actualidad) de un periodo aproximado de 7 años, sobre ensayos o estudios clínicos que contienen “lung cáncer” (cáncer de pulmón) referenciado en el artículo, obteniendo un resultado de 66.190 entradas a artículos que hacen referencia a este tipo de cáncer, en contraste con el resultado de 52.575 entradas en un periodo de 20 años (comprendido entre el 1990 y 2009) (Consulta realizada en https://www.ncbi.nlm.nih. gov/pubmed/?term=%22LUNG + CANCER%22 en marzo de 2018).

Una de las áreas importantes de investigación sobre este tipo de cáncer está focalizada en el concepto de que las mutaciones oncogénicas afectan a los sistemas de señalización que rigen el destino celular, dotando a las células tumorales de algunas características que hacen que su comportamiento se identifique como maligno. Unas de las principales vías estudiadas dentro de los sistemas de señalización celular, son las MAPK (Mitogen-activated protein Kinases) que regulan una variedad de actividades celulares como la

proliferación, la diferenciación, la supervivencia y la muerte (apoptosis) (Kyung Kim & Choi, 2010). En esta línea, ya hace más de 30 años se identificaron que existían varios tipos de tumores que presentaban mutaciones sobre miembros de la familia de proteínas RAS (K-RAS, N-RAS y H-RAS), por lo que las proteínas codificadas por estos genes pasaban a ser objetivos de interés para el tratamiento del cáncer. Se identificaba a K-RAS como la isoforma más comúnmente mutada con más del 20% de todos los cánceres humanos, respecto al N-RAS con el 8% y H-RAS con 3,3% de los cánceres (Baines, Xu, Dapeng, & Der, Channing J, 2011). Concretamente en las vías de señalización MAPK/ERK, es donde se han identificado muchas de las mutaciones asociadas al cáncer en la familia de proteínas RAS y concretamente mutaciones en el gen K-RAS relacionadas con el cáncer de pulmón y colón (Kyung Kim & Choi, 2010).

Por último y no por ello menos importante, cabe destacar como otro elemento de justificación del trabajo, el hecho de que en mi ámbito actual de actuación profesional, Centro Médico Teknon – Grupo Quironsalud, él área de investigación oncológica y en concreto el área de Pulmón, tienen una relevancia destacada con varias líneas de investigación abiertas referentes al cáncer de pulmón, que pueden facilitar el acceso a datos actualizados de ensayos o estudios clínicos, permitiendo poder realizar un proceso de análisis aplicando herramientas de bioinformática y bioestadística, para contrastar los resultados obtenidos con los obtenidos en el trabajo original. Esta comparación puede aportar nuevos enfoques al resultado final original.

1.2 Objetivos del Trabajo

Para la consecución del presente trabajo se han identificado y establecido tres objetivos estratégicos globales, que marcan las líneas generales de desarrollo del trabajo y que se concretan en diferentes objetivos más específicos. A continuación se enumeran tanto los objetivos globales estratégicos, como los específicos de cada uno.

1.2.1 Objetivos generales

A continuación se detallan los tres objetivos estratégicos del proyecto que se han planteado para la consecución óptima del trabajo y que son:

1- Identificar la implicación de la mutación gen K-RAS en la oncogénesis del cáncer a través de las disfunciones celulares que genera.
2- Determinar la incidencia del cáncer de pulmón causado por la mutación K-RAS versus otros tipos de cáncer.
3- Contrastar resultados de un estudio realizado sobre la efectividad de algún fármaco para el tratamiento de la mutación del gen K-RAS o bien, su implicación en la supervivencia de pacientes.

Aunque no identificado explícitamente como objetivo en el documento, pero no por ello no establecido, se ha marcado un objetivo global del trabajo en referencia a intentar alcanzar en nivel más alto posible de calidad del trabajo, basado tanto en la orientación, la metodología empleada, el contenido y la memoria final obtenida, así como la presentación realizada. Como objetivo básico se plantea obtener la excelencia de dicho proceso.

1.2.2 Objetivos específicos

Cada uno de los objetivos estratégicos globales planteados se materializa en diferentes objetivos más específicos que se detallan a continuación. El primer dígito de la enumeración de estos objetivos específicos hace referencia al objetivo global al que se relaciona.

1.1- Realizar un análisis global de las vías MAP Kinasas y específico de la vía MAPK/ERK.

1.2- Identificar la implicación de la familia de genes RAS en la oncogénesis de algunos tipos de cáncer.

1.3- Identificar y describir la mutación del gen K-RAS y las funciones celulares alteradas.

1.4- Identificar diferentes dianas terapéuticas relacionadas con la mutación del gen K-RAS

1.5- Identificar fármacos probados para el tratamiento de los efectos del gen K-RAS mutado y sus resultados.

2.1- Obtener datos estadísticos sobre la incidencia en la población de diferentes tipos de cáncer relacionados con mutaciones génicas en la vía MAPK/ERK en la familia de proteínas RAS.

2.2- Realizar análisis poblacional comparativo con los datos estadísticos, respecto al cáncer de pulmón.

3.1- Realizar un análisis bioestadístico con los datos obtenidos del estudio clínico actual, sobre muestras tumorales de adenocarcinoma de pulmón con mutación de KRAS. Verificación de la existencia de relación entre el nivel de expresión de genes afectados por mutación de KRAS en el cáncer de pulmón y la supervivencia de los pacientes.

3.2- Realizar un análisis comparativo de resultados, identificando diferencias, si las hay, entre el análisis bioestadístico realizado en el trabajo y los resultados obtenidos por el estudio original.

1.3 Enfoque y método seguido

Este proyecto se plantea en el ámbito de la investigación médica del cáncer de pulmón, enfocado a la identificación de dianas terapéuticas efectivas para su tratamiento y remisión, logrando aumentar, en lo posible, la esperanza de vida de los pacientes con esta patología.

Se pueden diferenciar dos partes fundamentales en el trabajo: una primera parte de investigación teórica sobre el funcionamiento e implicación de las vías de señalización o rutas de transducción de señales celulares en los seres vivos eucariotas, en la oncogénesis del cáncer de pulmón; y una segunda parte más práctica, de realización de un análisis biostadístico con datos de un estudio sobre el tratamiento de la desregularización del gen K-RAS y su implicación en el cáncer de pulmón.

Para la realización del proceso de investigación sobre las vías MAP kinasas, se aplicará una estrategia de análisis Top-Down, partiendo de la identificación más global de estas vías y profundizando en el análisis de forma más particular hasta

llegar a la identificación de la mutación del gen K-RAS en la vía ERK como factor iniciador de la carcinogénesis del cáncer de pulmón. En la figura F 1-1 se muestra gráficamente la estrategia Top-Down implementada para el análisis y desarrollo de los diferentes objetivos del proyecto.

Esta estrategia de análisis permite realizar el proceso de análisis e investigación de forma ordenada y completa para la consecución de los objetivos planteados. Permite ir enfocando cada vez con mayor profundidad las diferentes áreas del proyecto, partiendo del ámbito más global de la señalización celular, con el proceso de estudio y definición global de las vías de señalización MAPK y profundizando o bajando ("down") en el análisis más específico de las vías de interés del proyecto, la MAPK/ERK y de los agentes que tienen una participación relevante en el proceso de oncogénesis. Esta estrategia permite no perder el enfoque del proyecto durante su desarrollo, centrando la atención en las partes relevantes del mismo y aprovechando la información global obtenida en la etapa anterior, centrando cada vez más la atención en el eje central del análisis.

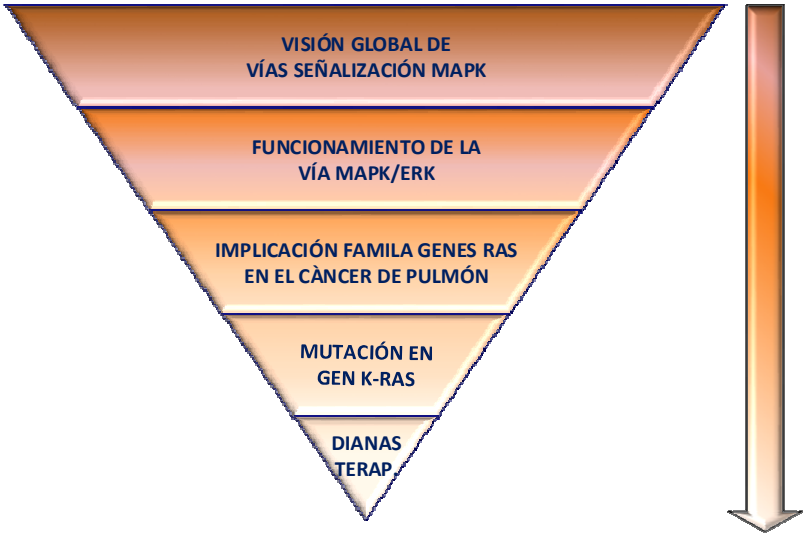


Figura F 1.1. *Estrategia de análisis "Top-Down", parte investigación teórica.*

Como continuación del proceso teórico del trabajo, en referencia al estudio comparativo de la incidencia de diferentes tipos de cáncer, con el cáncer de pulmón y partiendo del principio en que se basa la conocida frase "una imagen vale más que mil palabras" (anónimo), se aplicará una metodología fundamentalmente gráfica, basada en diferentes tipos de gráficos adecuados a las diversas características de los datos a mostrar: gráficos histogramas, gráficos de caja, gráficos de sectores, gráficos de puntos, gráficos de barras, etc. Como herramienta informática para la representación, tanto numérica como gráfica de los datos, se utilizará el paquete de desarrollo integrado RStudio (Version 1.1.383 – © 2009-2017 RStudio, Inc.) orientado a la computación estadística y gráfica, que contiene el entorno de programación de R (versión 3.3.0) (© The R Foundation), ampliamente utilizado por la comunidad científica.

Para el proceso práctico de análisis, con la finalidad de verificar la existencia de una relación entre el nivel de expresión de genes afectados por mutación de KRAS en el cáncer de pulmón y la supervivencia de los pacientes, se tratarán y analizarán los datos obtenidos del estudio clínico realizado en la Tesis doctoral

con el título: “Carcinoma de pulmón no célula pequeña, con mutación del gen KRAS…” (Gil Moreno, 2018) de pacientes con cáncer de pulmón. Al igual que en el proceso de estudio gráfico estadístico, se ha utilizado el paquete de desarrollo integrado RStudio (Version 1.1.383 – © 2009-2017 RStudio, Inc.), dado su alto rendimiento y su alta capacidad de análisis en el ámbito de la Bioinformática y Bioestadística. En adición a estas altas prestaciones para el análisis estadístico, la gran cantidad de funciones estadísticas disponibles en el entorno de programación de R, que son usadas ampliamente por la comunidad de investigación científica y en la mayoría de los casos son de libre difusión y uso, son el complemento idóneo para abordar con éxito un proyecto de análisis bioestadístico de estas características.

El esquema de trabajo planteado inicialmente para el análisis bioestadístico de este estudio ha sido el siguiente:

- Proceso de carga y exploración de datos disponibles del estudio. Exploración numérica y gráfica.
- En caso de necesidad, se efectuará un proceso de normalización de los datos disponibles, para su posterior análisis estadístico.
- Definición del modelo estadístico adecuado para la realización de la verificación de la hipótesis planteada en el experimento (pipeline de trabajo).
    - Este punto se deberá desarrollar en más profundidad una vez se conozcan con detalle los datos disponibles del estudio y la hipótesis experimental a verificar.
- Desarrollo y ejecución del modelo aplicando las funciones requeridas en R (© The R Foundation) y mostrando los resultados numéricos y gráficos obtenidos.
- Evaluación de resultados obtenidos.
- Comparativa de resultados obtenidos con los resultados originales, identificación de diferencias y posibles líneas posteriores de estudio.

Para poder efectuar el informe final del análisis estadístico realizado con los datos obtenidos del estudio, se utilizará el lenguaje de marcado Markdown dentro de la plataforma de RStudio (FOAS: Foundation for Open Access Statistics - R Consortium). Este tipo de lenguaje de marcado utilizado dentro del entorno de programación de RStudio, permite la coexistencia de las partes de código de R, donde se apliquen las funciones estadísticas requeridas para el análisis, con el texto explicativo introducido para la elaboración de los resultados obtenidos. Este entorno de trabajo, permite al investigador compartir los resultados junto con el código realizado, para que los resultados puedan ser reproducidos en cualquier momento por otros investigadores, aprovechando así los resultados obtenidos en el trabajo para complementarlos con nuevos elementos de análisis o ampliando el ámbito de trabajo. Por otra parte, RMarkdown (© Copyright 2016 RStudio Inc.) permite la generación de un informe de salida en diferentes formatos (Word, PDF o HTML), que facilitan la compartición de forma rápida de los resultados reproducibles con la comunidad científica.

Como se indica en el apartado 1.5.3, como uno de los entregables que se aportan en el TFM para el apartado del análisis comparativo Bioinformático, se incluirán el fichero reproducible de RStudio, el informe resultante de su ejecución

en formato PDF y el mismo informe de resultado pero en formato HTML, para que este pueda ser presentado en un entorno Web y consultado por Internet.

1.4 Planificación del Trabajo

Para la realización del trabajo, se han desglosado cada uno de los objetivos específicos definidos en el apartado 1.2.2 en las tareas principales a realizar para la consecución de los mismos, que se detallan en el apartado 1.4.1. En el apartado 1.4.2 se establece la planificación del trabajo completo, teniendo en consideración la estimación de todas las dependencias existentes entre cada una de las tareas a desarrollar de forma correcta y coherente, con el objetivo de la consecución de cada hito definido según la planificación estimada.

Adicionalmente, en el apartado 1.4.3, se detallan los hitos de control que se han establecido para asegurar la correcta evolución del desarrollo del trabajo.

1.4.1 Tareas

Para cada uno de los objetivos globales y específicos marcados en el trabajo, se han identificado a un nivel más concreto las tareas requeridas que permitan alcanzar dichos objetivos con el nivel de detalle necesario para elaborar la memoria final del trabajo con el nivel de calidad más alto posible. Cada una de las tareas está identificada numéricamente, relacionada con el objetivo al que hace referencia, donde el primer dígito de identificación corresponde al objetivo global al que pertenece.

Se han identificado adicionalmente las tareas relacionadas con el proceso global de desarrollo y entrega del trabajo, que a su vez están relacionados con la definición de conclusiones, la elaboración de la memoria, la preparación de la defensa pública y con la confección de la presentación.

Para una mejor organización y planificación, las tareas se han agrupado en cuatro bloques de trabajo: del A al D. Cada bloque contiene las tareas de una línea concreta de trabajo a las que hace referencia:

A. Analisis global top-down de las vías de señalización celular MAPK, implicación de la mutación del gen K-RAS en la oncogénesis.
    - 1.1- Búsqueda información sobre las vías MAP Kinasas globalmente y sobre el funcionamiento de la vía MAPK/ERK de forma específica.
    - 1.2- Describir el funcionamiento de la vía MAPK/ERK, identificando las funciones de la familia de genes RAS en esta vía y su implicación en la oncogénesis del cáncer.
    - 1.3- Buscar, identificar y describir la mutación del gen K-RAS y las funciones celulares alteradas.
    - 1.4- Buscar e identificar diferentes dianas terapéuticas relacionadas con la mutación del gen K-RAS, así como fármacos probados para el tratamiento de las funciones alteradas y sus resultados.
B. Estudio poblacional de la incidencia del cáncer de pulmón con implicación de la mutación en genes de la familia RAS en comparación a otros tipos de cáncer.

2.1- Realizar una búsqueda de datos estadísticos sobre incidencias en la población de diferentes tipos de cáncer relacionados con mutaciones génicas en la vía MAPK/ERK en la familia de proteínas RAS.

2.2- Generar gráficas estadísticas comparativas de análisis poblacional a partir de los datos obtenidos de los diferentes tipos de cáncer en la tarea 2.1.

C. Analisis bioestadístico comparativo de un estudio clínico actual perteneciente a la Tesis doctoral con el título: "C*arcinoma de pulmón no célula pequeña, con mutación del gen KRAS…"* (Gil Moreno, 2018) de pacientes con cáncer de pulmón.

3.1- Obtención de los datos del estudio original.

3.2- Exploración y análisis de los datos disponibles para el estudio.

3.3- Revisión y ajuste del diseño del *pipeline* del análisis bioestadístico a realizar en función de los datos obtenidos del estudio original.

3.4- Proceso de análisis bioestadístico de los datos obtenidos

3.5- Realización de comparación de resultados, extracción de aspectos diferenciales respecto al estudio original y obtención de conclusiones.

D. Elaboración de memoria, presentación y defensa pública.

4.1- Extracción de conclusiones del TFM.

4.2- Elaboración de la memoria del TFM.

4.3- Confección de la presentación y preparación de la defensa pública.

1.4.2 Calendario

Como parte fundamental del proyecto, se ha definido una planificación temporal para el desarrollo completo del trabajo. Siguiendo tanto las pautas indicadas en el Plan Docente, como de las tareas, subtareas e hitos identificados en el apartado 1.4.1, se ha distribuido el total de horas de dedicación estimadas para su desarrollo en función de la complejidad estimada de cada tarea, a fin y efecto de controlar su correcto desarrollo sin desviaciones que puedan poner en peligro la consecución de los objetivos establecidos en los plazos señalados por el Plan Docente del TFM.

Como herramienta de planificación se ha elaborado un diagrama de Gantt, que se muestra en la figura F 1.2 y F 1.3. En este diagrama se muestran, de forma estructurada, las diferentes tareas a realizar organizadas según el esquema de bloques de trabajo indicados en el apartado de tareas, indicando el número identificador de cada una, y agrupadas según la estructura marcada en el Plan Docente. Para cada subtarea se indica la descripción de la misma, la duración estimada en días para su desarrollo, así como la referencia identificadora de cada una, que indirectamente nos da la información sobre con qué objetivo global está relacionada.

El calendario se muestra en semanas por mes. Para la realización del trabajo reflejado en el diagrama de Gantt (figuras F 1.2/3) se han considerados como jornada lectiva, todos los días de la semana, de lunes a domingo, sin excluir festivos, dado que la disponibilidad horaria de lunes a viernes solo es de 4 horas diarias, por lo que se hace necesario incluir como lectivos los fines de semana y festivos, para poder tener la dedicación requerida para finalizar el trabajo en los plazos requeridos. Basados en esta planificación, se estiman 101 días de trabajo, que con una dedicación de 4 horas por día representan un total de 404 horas de dedicación. Las horas de dedicación en los fines de semana y festivos, podrá variar en función de las posibilidades, pero en cualquier caso la dedicación final del fin de semana será siempre la estimada.

Los hitos establecidos en la planificación están indicados con un punto en rojo en el calendario, indicando la entrega o control que se debe realizar en la fecha indicada.

1.4.3 Hitos

Como se apuntaba en la introducción del apartado, se han definido una serie de hitos de control para validar la correcta evolución y desarrollo del trabajo, con la finalidad de detectar posibles desviaciones y efectuar las correcciones que se estimen oportunas para no alterar ni la planificación, ni el contenido del trabajo. Se han establecido como hitos base, los entregables marcados por el Plan Docente del TFM (como entregas de PEC) y adicionalmente hitos intermedios de control, para verificar desviaciones en la planificación, así como los elementos de riesgo identificados en el apartado 4.4, que se hayan podido producir y por tanto a los que se les debe aplicar el Plan de contingencias definido para cada riesgo identificado.

En el calendario mostrado en el apartado 1.4.2, se han identificado explícitamente cada uno de estos hitos, en las fechas planificadas y con el control o entregable asociado a él.

Los hitos establecidos son:

- *Hito 1. Plan de Trabajo.* Fecha: 19-03-2018
  - Entrega de documento PEC1.
- *Hito 2. Control desvío planificación 1.* Fecha: 08-04-2018
  - Primer control de revisión de planificación, para detectar desvíos en los plazos indicados y aplicación de plan de contingencias si se requiere.
- *Hito 3. Desarrollo TFM Fase 1.* Fecha: 23-04-2018
  - Entrega de documento PEC2 y revisión de planificación.
- *Hito 4. Control desvío planificación 2.* Fecha: 07-05-2018
  - Segundo control de revisión de planificación, para detectar desvíos en los plazos indicados y aplicación de plan de contingencias si se requiere.

Tarea	Nombre de tarea	Duración	Comienzo	Fin
	Trabajo Final de Máster	**101 días?**	**lun 05/03/18**	**mié 13/06/18**
	Hito 0. Propuesta TFM (PEC0)	1 día?	lun 05/03/18	lun 05/03/18
Plan TFM	**Elaboración plan de trabajo TFM**	13 días	mar 06/03/18	dom 18/03/18
HITO 1	Hito 1. Plan de Trabajo (PEC1)	1 día?	lun 19/03/18	lun 19/03/18
	Desarrollo del TFM Fase 1	**35 días**	**mar 20/03/18**	**lun 23/04/18**
A	**A. Analisis global top-down de las vías de señalización celular MAPK, implicación de la mutación del gen K-RAS en la oncogénesis.**	**21 días**	**mar 20/03/18**	**lun 09/04/18**
1.1	1.1- Búsqueda información sobre las vías MAP Kinasas globalmente y sobre el funcionamiento de la vía MAPK/ERK de forma específica.	3 días	mar 20/03/18	jue 22/03/18
1.2	1.2- Describir el funcionamiento de la vía MAPK/ERK, identificando las funciones de la familia de genes RAS en esta vía y su implicación en la oncogénesis del cáncer.	4 días	vie 23/03/18	lun 26/03/18
1.3	1.3- Buscar, identificar y describir la mutación del gen K-RAS y las funciones celulares alteradas.	9 días	mar 27/03/18	mié 04/04/18
1.4	1.4- Buscar e identificar diferente dianas terapéuticas relacionadas con la mutación del gen K-RAS, así como fármacos probados para el tratamiento de las funciones alteradas y sus resultados.	5 días	jue 05/04/18	lun 09/04/18
HITO 2	Hito 2. Control desvio planificación	1 día	dom 08/04/18	dom 08/04/18
B	**B. Estudio poblacional de la incidencia del cáncer de pulmón con implicación de la mutación en genes de la familia RAS en comparación a otros tipos de cáncer.**	**14 días**	**mar 10/04/18**	**lun 23/04/18**
2.1	2.1- Realizar una búsqueda de datos estadísticos sobre incidencias en la población de diferentes tipos de cáncer relacionados con mutaciones génicas en la vía MAPK/ERK en la familia de proteínas RAS	5 días	mar 10/04/18	sáb 14/04/18
2.1	2.2- Generar gráficas estadísticas comparativas de análisis poblacional a partir de los datos obtenidos de los diferentes tipos de cáncer en la tarea 2.1.	8 días	dom 15/04/18	dom 22/04/18
HITO 3	Hito 3. Desarrollo TFM Fase I (PEC2)	1 día	lun 23/04/18	lun 23/04/18
	Desarrollo del TFM Fase 2	**28 días**	**mar 24/04/18**	**lun 21/05/18**
C	**C. Analisis bioestadístico comparativo de un estudio clínico actual de pacientes carcinoma de pulmón no célula pequeña, con mutación del gen KRAS .**	**27 días**	**mar 24/04/18**	**dom 20/05/18**
3.1	3.1- Obtención de los datos del estudio original.	1 día	mar 24/04/18	mar 24/04/18
3.2	3.2- Exploración y análisis de los datos disponibles para el estudio.	3 días	mié 25/04/18	vie 27/04/18
3.3	3.3- Revisión y ajuste del diseño de pipeline del análisis bioestadístico a realizar en función de los datos obtenidos del estudio original.	2 días	sáb 28/04/18	dom 29/04/18
3.4	3.4- Proceso de análisis bioestadístico de los datos obtenidos	16 días	lun 30/04/18	mar 15/05/18
HITO 4	Hito 4. Control desvio planificación II	1 día	lun 07/05/18	lun 07/05/18
3.5	3.5- Realización de comparación de resultados, extracción de aspectos diferenciales respecto al estudio original y obtención	5 días	mié 16/05/18	dom 20/05/18
HITO 5	Hito 5. Desarrollo TFM Fase 2 (PEC3)	1 día	lun 21/05/18	lun 21/05/18
	Redacción Memoria y presentación	**24 días**	**lun 21/05/18**	**mié 13/06/18**
D	**D. Elaboración de memoria, presentación y defensa pública.**	**23 días**	**lun 21/05/18**	**mar 12/06/18**
4.1	4.1- Extracción de conclusiones del TFM.	4 días	lun 21/05/18	jue 24/05/18
4.2	4.2- Elaboración de la memoria del TFM.	14 días	lun 21/05/18	dom 03/06/18
HITO 6	Hito 6. Entrega memoria (PEC4)	1 día	lun 04/06/18	lun 04/06/18
4.3	4.3- Confección de la presentación y preparación de la defensa pública.	4 días	sáb 09/06/18	mar 12/06/18
HITO7	Hito 7. Entrega presentación (PEC5a)	1 día	mié 13/06/18	mié 13/06/18

***Figura F 1.2**. Planificación de tareas del TFM.*

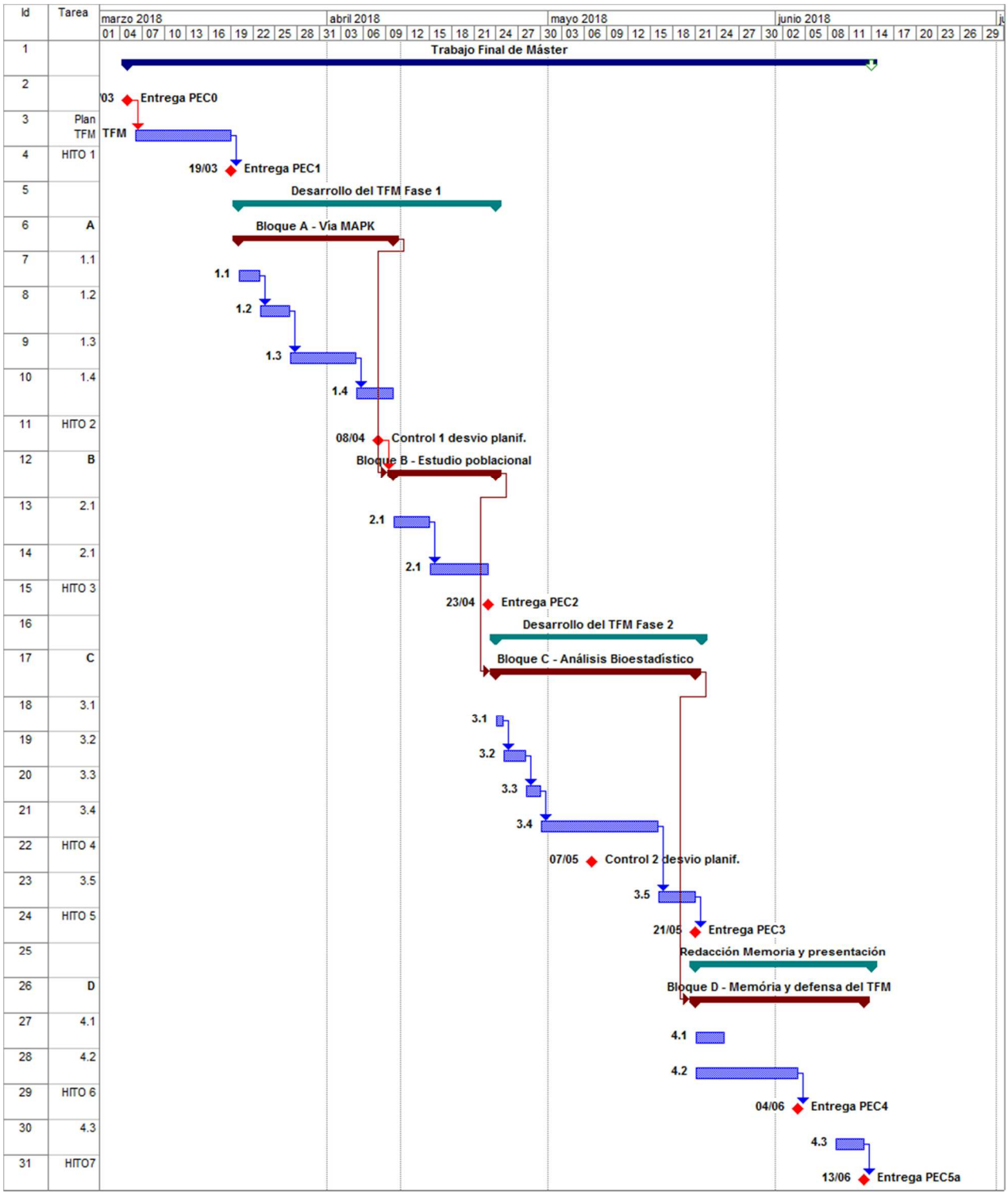


Figura F 1.3. *Diagrama de Gantt del TFM.*

- *Hito 5. Desarrollo TFM Fase 2.* Fecha: 21-05-2018
  - Entrega de documento PEC3 y revisión de planificación.
- *Hito 6. Entrega memoria de TFM.* Fecha: 04-06-2018
  - Entrega de documento PEC4.
- *Hito 7. Entrega presentación (PEC5a).* Fecha: 04-06-2018

- Entrega de documento PEC5a.

1.4.4 Análisis de riesgos

A parte del riesgo implícito en todo trabajo, como es el factor tiempo, debido a una estimación incorrecta de la dedicación y/o incumplimiento de la dedicación, en base a las características y contenido del trabajo, se han identificado algunos factores adicionales de riesgo a considerar, los cuales se relacionan a continuación y sobre los que se establecen los planes de contingencia que se describen más adelante.

Los factores de riesgo adicionales identificados son:

1. Por desconocimiento de la existencia de la información pública requerida para el estudio. Dificultad en la búsqueda y localización de información estadística sobre la incidencia del cáncer, causado por mutaciones en la familia de genes RAS, para realizar el estudio poblacional comparativo entre el cáncer de pulmón, objetivo del trabajo, versus otros tipos de cáncer.
2. En el momento de elaborar el presente análisis de riesgos, no se dispone de suficiente información sobre el estudio clínico sobre el cual se pretende realizar el estudio Bioinformático descrito en el bloque C. Este desconocimiento es un riesgo en sí, dado que no se conoce aún lo suficiente sobre la tipología ni complejidad de los datos disponibles que serán la base de la tarea de análisis del trabajo, y por ello, se deben contemplar los siguientes riesgos:
    a. El *pipeline* de trabajo de análisis diseñado no se ajusta a la realidad del proceso que se deberá realizar.
    b. La complejidad de los análisis a realizar es mayor que la inicialmente estimada. El alcance supera al establecido en el TFM.
    c. En base a los dos factores anteriores, la planificación estimada para esta parte del trabajo no es correcta.

En el caso de producirse alguno de los factores de riesgo relacionados durante el desarrollo del trabajo, este debería identificarse, bien al inicio de la tarea correspondiente o en el punto de control establecido como hito en la planificación. Una vez identificada la existencia de alguno de los factores, se deberá aplicar el elemento corrector necesario, siguiendo el plan de contingencias establecidos para cada riesgo, minimizándolo o eliminándolo, si es posible, de forma que se pueda continuar el desarrollo del TFM con las correcciones introducidas y sin alterar los plazos finales de ejecución definidos.

En el caso de la detección de desviaciones en las planificaciones estimadas, se analizará la causa de las mismas. En caso de desviaciones por falta de horas de dedicación según la planificación, se deberán añadir horas adicionales a las previstas para los siguientes días, de forma que se pueda corregir la desviación identificada. En el caso de producirse la desviación por la imposibilidad de conseguir algún tipo de información relacionada con el contenido del trabajo, se solicitará apoyo al Director del TFM para intentar obtener la información requerida de otras fuentes a las que él pueda acceder. En referencia a los factores de riesgo adicionales, se identifican a continuación los planes de

contingencias para cada uno de ellos, siguiendo la misma numeración reflejada para cada factor de riesgo:

1. En caso de que no haya disponibilidad de datos públicos de estadísticas de la incidencia de cáncer relacionados con mutaciones génicas de la familia de RAS (contrastado con el Director del TFM), se enfocará el análisis comparativo poblacional en base a datos globales por tipos de cáncer en Europa y España, entre el cáncer de pulmón y otros tipos de cáncer con una incidencia poblacional alta.
2. Una vez se disponga de la información detallada del estudio clínico sobre el que se realizará el proceso de análisis bioinformático:
    a. Si el pipeline inicial no es adecuado para el proceso de análisis a realizar, se modificará éste y se ajustará a las necesidades, considerando el mantener la planificación y los objetivos globales del trabajo.
    b. Si la complejidad del proceso a realizar una vez conocidos los datos y el proceso realizado en el trabajo original es mayor que la establecida para el alcance del TFM, se modificará el trabajo a realizar adecuándolo al alcance real establecido para el TFM, sin perder por ello el punto de interés científico requerido para un trabajo de estas características, siempre de acuerdo con las indicaciones del Director del TFM.
    c. Si se ha producido alguno de los riesgos indicado en a y b, que afecten a la planificación, se deberán introducir los elementos requeridos en las tareas comprometidas por a y/o b, para que la planificación final no se vea alterada y se pueda conseguir el resultado esperado.

1.5 Resultados esperados

En este apartado se describen brevemente los resultados que se esperan obtener como resultado entregable del TFM, que se concretan en los siguientes cinco entregables:

1.5.1 Plan de trabajo

Se entregará en formato PDF el presente informe con el título “PLAN DE TRABAJO DEL TFM”, donde se describe el ámbito global del trabajo, su justificación, los objetivos planteados, las metodologías aplicadas para llevarlo a cabo, la planificación prevista, un análisis de riesgos básico, planes de contingencias previstos, así como los resultados esperados entregables con el trabajo y la estructura prevista del proyecto.

1.5.2 Memoria

Siguiendo las indicaciones del Plan Docente, se aportará como resultado principal del TFM una memoria escrita en formato PDF con todo el desarrollo del proyecto, según la estructura indicada en el apartado 1.6 de este informe.

1.5.3 Producto

Tal y como se apuntaba en el apartado de 1.3 de Enfoque y método a seguir de este documento, adicionalmente a la memoria del TFM, se entregará el resultado obtenido del bloque 3 del TFM, referido al proceso de análisis Bioestadístico realizado con los datos de un estudio clínico de pacientes con carcinoma de pulmón. Este entregable estará dividido en tres ficheros:

1. Fichero script de Rstudio - Rmarkdown (© Copyright 2016 RStudio Inc.) del proceso completo del análisis realizado, con la extensión ".Rmd"
2. Fichero de informe de los resultados obtenidos del proceso de análisis, en formato PDF.
3. Ficheros del informe de los resultados obtenidos del proceso de análisis en formato HTML, para su compartición en Web.

1.5.4 Presentación virtual

Se entregará un documento con formato Power Point o Prezzi, como presentación del TFM para el proceso de defensa pública. El formato final se determinará en función de las posibilidades y necesidades para la realización de la presentación, más adelante al finalizar el desarrollo del TFM.

1.5.5 Autoevaluación del proyecto

Durante el proceso de desarrollo del TFM, coincidiendo con los hitos establecidos en el calendario de planificación, se efectuará un seguimiento del cumplimiento de los objetivos temporales y de contenidos establecidos, que servirán de base para un proceso de autoevaluación objetiva sobre el desarrollo del TFM.

Al final del proceso de realización del TFM se cumplimentará el documento de autoevaluación entregado por el Director del proyecto a tal efecto, el cual se añadirá al conjunto de entregables del proyecto.

1.6 Estructuración de memoria del proyecto

Tal como se ha mostrado en el apartado de planificación (1.*4.2 Calendario*), la estructura del TFM está basada en las indicaciones descritas en el Plan Docente, donde se han identificado las diferentes entregas de control del trabajo.

A parte de los documentos definidos en las diferentes entregas de control, que se indican a continuación, se especifica la estructura del desarrollo del TFM, indicando la estructura de los contenidos en cada fase según las entregas a realizar indicadas en el *Plan Docente*.

La estructura del proyecto coincide con el contenido de la memoria del TFM y que se describe a continuación. Se han incluido los diferentes entregables definidos para el proyecto.

Trabajo Final de Máster

1 Introducción

En el apartado de Introducción se incluirá la información realizada en el primer entregable de Plan de Trabajo (PEC1), que contendrá los siguientes apartados coincidentes con los definidos en este entregable:

1.1 Contexto y justificación del Trabajo

1.2 Objetivos del Trabajo

1.3 Enfoque y método seguido

1.4 Planificación del Trabajo

1.5 Breve sumario de productos obtenidos

1.6 Breve descripción de los otros capítulos de la memoria

- **Entrega** del Plan de trabajo en formato de **PEC1**.

El cuerpo del TFM está estructurado en tres capítulos que conforman el núcleo principal del proyecto y que son los siguientes:

2 Vías de señalización celular MAP Kinasas y su relación con mutaciones en KRAS

2.1 Vías MAP Kinasas

2.2 Funcionamiento de la vía MAPK/ERK: Familia KRAS y su implicación oncogénica

2.3 Mutaciones en KRAS: subtipos según mutación

2.4 Dianas terapéuticas relacionadas con KRAS

3 Estudio poblacional de incidencia del cáncer de pulmón

3.1 Estadística global de incidencia de Cáncer en el mundo y en Europa

3.2 Incidencia en la población de cánceres relacionados con MAPK/ERK

3.3 Comparativa de la incidencia del cáncer de pulmón con otros tipos de cáncer

- **Entrega** del Desarrollo TFM Fase 1 en formato de **PEC2.**

4 Estudio Bioestadístico comparativo:

Relación entre supervivencia y nivel de expresión de genes con mutación de KRAS en el cáncer de pulmón

4.1 Descripción del estudio original

4.2 Análisis bioinformático y bioestadístico

4.2.1 Objetivos del proceso de análisis

4.2.2 Pipeline para el proceso de análisis

4.2.3 Proceso de análisis

4.2.4 Conclusiones del análisis de supervivencia

4.2.5 Comparativa de resultados con estudio original

Los subapartados adicionales de los capítulos 2,3 y 4 se detallarán durante el desarrollo del TFM.

El objeto principal del proceso de análisis del capítulo 4 queda supeditado a los datos que finalmente se puedan disponer del estudio original para efectuar el proceso de análisis y que podrá variar al indicado en este momento.

- **Entrega** del Desarrollo TFM Fase 2 en formato de **PEC3.**

5 Conclusiones

En este apartado se incluirán las conclusiones globales extraídas del TFM respecto al estudio global realizado sobre el cáncer de pulmón relacionado a las mutaciones en KRAS.

6 Glosario

7 Bibliografía

8 Anexos

- **Entrega** de la memoria del TFM en formato de **PEC4.**

Capítulo 2

2. Vías de señalización celular.
Vías MAPK: Relación con mutaciones en KRAS

Las vías de señalización celular proporcionan los mecanismos de organización molecular a las células para adaptarse al entorno según los estímulos externos que reciben. La capacidad de las células para identificar la información extracelular que reciben, traducirla, interactuar con otras moléculas y responder biológicamente con un patrón fenotípico concreto, viene determinada por el proceso de transducción de las señales externas recibidas a través de las vías o rutas de señalización intracelular. Estas vías o rutas de señalización regulan múltiples funciones celulares, especialmente su expresión génica, así como su inhibición (Victor. M. Valdespino-Gómez et al., 2015).

Las moléculas que inician el proceso de transducción de la señal extracelular en una vía de señalización intracelular son diferentes tipos de proteínas que actúan como receptores de la superficie celular, que reciben una molécula de señal externa, generalmente a nivel de la membrana plasmática, activando así una vía de señalización intracelular que está mediada por una serie de moléculas transductoras internas (proteínas de señalización), que a su vez pueden interactuar con una proteína diana, alterando esta para que modifique el comportamiento de la célula (Victor. M. Valdespino-Gómez et al., 2015; Bruce Alberts, 2002).

Las proteínas receptoras se pueden clasificar principalmente en tres tipologías: receptores asociados a canales iónicos, receptores asociados a proteínas G (monoméricos y heteotriméticos) y los receptores asociados con enzimas, habitualmente con actividad Kinasa, que participan en el control de comportamientos celulares complejos (Victor. M. Valdespino-Gómez et al., 2015).

La integración de múltiples vías de señalización intracelular, en una forma adecuada, regula los procesos complejos de la célula, incluida la proliferación, la diferenciación, la respuesta al estrés, la apoptosis. La alteración en la regulación de estas vías de señalización puede conducir a comportamientos no deseables, generando una activación o desactivación de la respuesta celular no adecuada, origen de muchas enfermedades y en particular de muchos tipos de cáncer (Victor. M. Valdespino-Gómez et al., 2015).

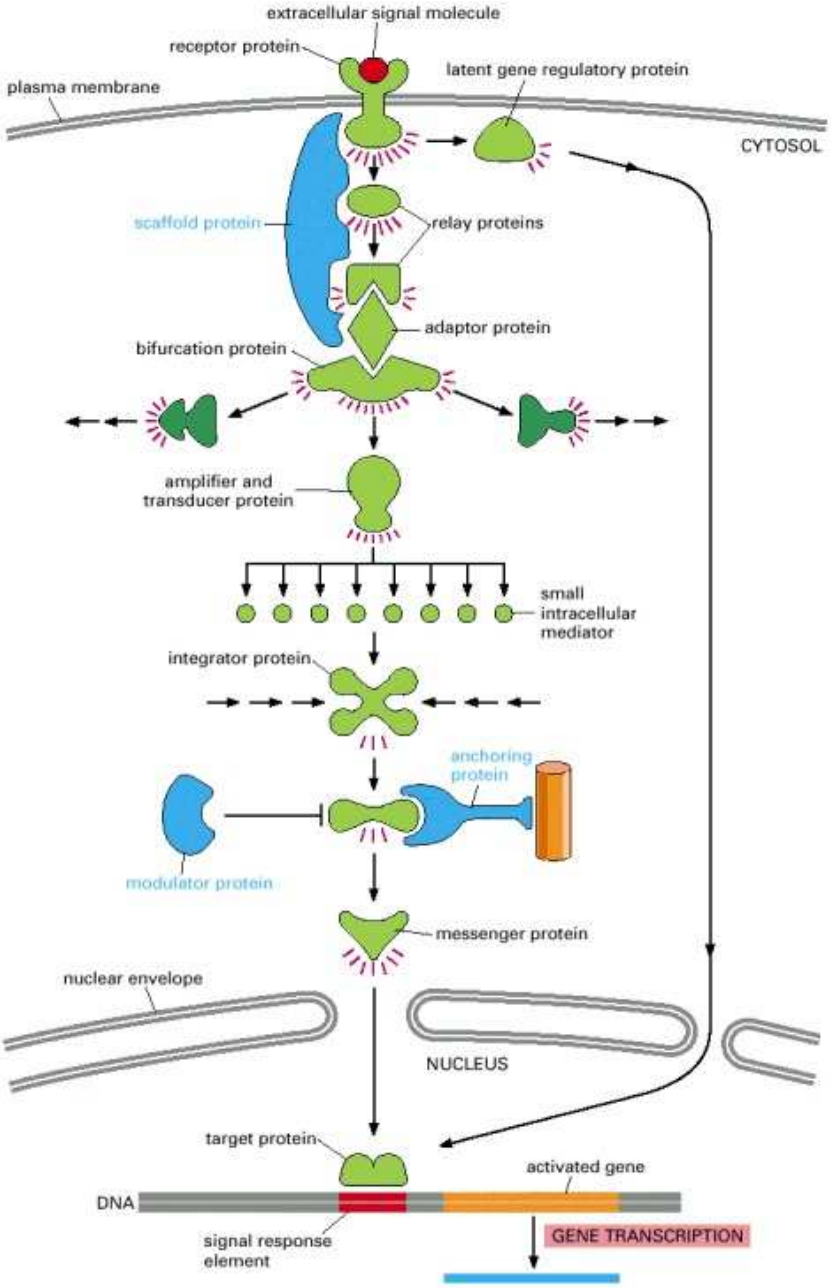


Figura F 2.1. *Proteínas en vía de señalización.*
Diferentes tipos de proteínas que pueden actuar en una vía de señalización. La activación de una vía se establece a partir de la interacción directa de proteína-proteína, o por sus modificaciones covalentes como la fosforilación y acetilación. En la figura se muestra gráficamente el mecanismo general de funcionamiento de una vía de señalización (Bruce Alberts, 2002).

Las vías de señalización intracelular que participan en la regulación de los principales procesos celulares se podrían clasificar en dos grandes grupos: por una parte, un primer grupo con aquellas vías que emplean moléculas mensajeras pequeñas (segundos mensajeros) para la transmisión y amplificación del mensaje y por otra parte, un segundo grupo que esencialmente emplean proteínas como mensajeros externos, que entre otras está formado por las vías de las kinasas activadas por mitógenos, conocidas por *MAPK* (Victor. M.

Valdespino-Gómez et al., 2015). En los siguientes apartados se revisan globalmente estas vías y con más detalle la via MAPK/ERK.

2.1 Vías MAP Kinasas

Las rutas de proteínas Kinasas activadas por mitógenos (MAPK) están muy conservadas en Eucariotas, regulando varias actividades celulares, incluida la supervivencia, la proliferación, la diferenciación y la apoptosis o muerte (KEGG, Consultado en 2018). Se activan por una gran variedad de estímulos, fosforilan numerosas proteínas, incluyendo factores de transcripción, proteínas del citoesqueleto, kinasas y otras enzimas, que actúan sobre receptores tirosin kinasas (RTK). Los receptores RTK se encuentran acoplados a proteínas G (GPCR), como puede ser por ejemplo RAS, receptores de citoquinas o receptores serina/treonina kinasa (Maosong Qi & Elaine A. Elion, 2005; Gil Moreno, 2018).

Se ha comprobado que alteraciones en la regulación de la señalización MAPK están implicadas en el desarrollo de muchas enfermedades que afectan al cerebro, por ejemplo la enfermedad de Alzheimer, la enfermedad de Parkinson, esclerosis lateral amiotrófica, y en diferentes tipos de cáncer (Kyung Kim & Choi, 2010).

En los mamíferos se expresan cuatro rutas o cascadas principales de MAPK, que presentan un módulo de triple kinasa (MKKK-MKK-MAPK), que consiste en una MAP-kinasa- kinasa-kinasa (MKKK), que activa mediante fosforilación una MAP-kinasa-kinasa (MKK), que a su vez mediante fosforilación doble de los residuos de Thr (Treonina) y Tyr (Tirosina) puede activar por fosforilación a sus sustratos en residuos de Ser (Serina) o Thr (Treonina). En el núcleo los sustratos fosforilan factores de transcripción y regulan la expresión génica (Shen-Hsi Yang, 2013; KEGG, Consultado en 2018; Gil Moreno, 2018).

Dentro de las cascadas MAPK se encuentran actores intermedios que proporcionan diferentes mecanismos para detectar las entradas de otras vías de señalización para mejorar o suprimir la señal a la MAP kinasa (*Hubs*), que también puede amplificar la señal que reciben. Esta amplificación puede ocurrir si cada proteína sucesiva en la cascada es más abundante que su regulador, generando con ello que la señal de origen sea amplificada (Pearson, 2001).

2.1.1 Vías de señalización MAPK

En la Enciclopedia de Kioto sobre genes y genomas (KEGG: Kyoto Encyclopedia of Genes and Genomes, 2018) se recogen más de 400 vías de señalización celular. En la figura F 2.2, se muestra el mapa de las rutas de señalización MAPK confeccionado por KEGG. Enmarcado en recuadro rojo, se centra la atención en la ruta MAPK/ERK (KEGG, Consultado en 2018).

Las principales rutas o cascadas MAPK son las siguientes:

- **ERK1/2**, kinasa regulada por señal extracelular.

  En referencia a su activación, la cascada ERK1/2 generalmente responde a señales de factores de crecimiento, como puede ser el factor de crecimiento derivado de plaquetas, PDGF, el factor de crecimiento epidérmico, EGF, y el

factor de crecimiento nervioso, NGF, también responde a la insulina (Cargnello, Marie & Roux, Philippe P, 2011).

- **ERK5, kinasa regulada por señal extracelular 5, también llamada BMK1 (big MAP kinasa-1).**

  La ruta ERK5 se activa y aumenta su respuesta a factores de crecimiento, suero, estrés oxidativo e hiperosmolaridad (Cargnello, Marie & Roux, Philippe P, 2011).

- **JNK1/2/3,** cJun anino-terminal kinasa.

  Las cascadas JNK se activan en respuesta a situaciones de estrés celular, como son el choque térmico, radiación ionizante, estrés oxidativo, agentes que dañan el ADN, citoquinas, radiación UV, inhibidores de la síntesis de ADN y proteínas y privación del factor de crecimiento, también se pueden activar, aunque en menor medida, por factores de crecimiento, algunos ligandos de GPCR y suero (Cargnello, Marie & Roux, Philippe P, 2011).

- **Isoenzimas p38 kinasas**, proteínas p38α, p38β, p38γ y p38δ.

  Las cascadas p38, al igual que las JNK, se activan por algunas tensiones ambientales y citoquinas inflamatorias, como son el estrés oxidativo, irradiación UV, hipoxia, isquemia, interleucina-1 (IL-1) y factor de necrosis tumoral alfa (TNF-α) (Cargnello, Marie & Roux, Philippe P, 2011).

Adicionalmente, se encuentran las MAPK atípicas que tienen particularidades especiales, que comprenden ERK3/4, ERK7 y NLK (Nemo-like kinase) (Cargnello, Marie & Roux, Philippe P, 2011).

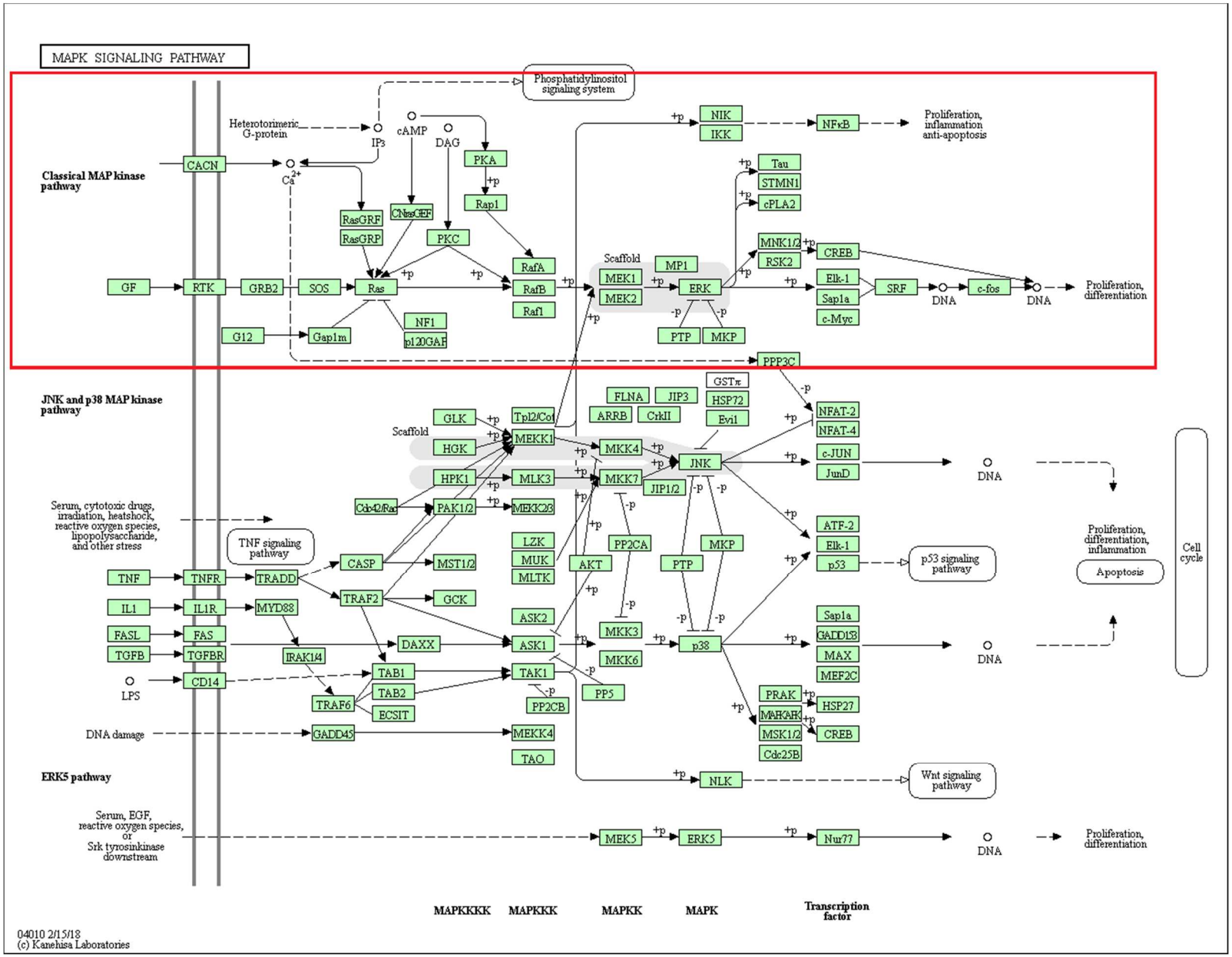


Figura F 2.2. *(KEGG, Consultado en 2018) Vías de señalización celular MAPK (hsa04010 MAPK signaling pathway)*

2.2 Funcionamiento de la vía MAPK/ERK: Familia RAS y su implicación oncogénica

2.2.1 Vía o cascada RAS-ERK

La vía de señalización MAPK/ERK o también conocida como RAS-ERK (RAS-RAF-MEK-ERK) es una de las vías más estudiadas durante los últimos años. Es una ruta de señalización muy conservada evolutivamente, que se activa por diferentes agentes o factores extracelulares, de aquí su nombre ERK (Extracellular signal-Regulated Kinases). Mediante receptores ubicados en la superficie celular trasmite las señales aguas abajo en la vía, para finalmente promover los procesos de proliferación y supervivencia celular. ERK está controlada por ciclos de retroalimentación clasificados en diferentes niveles, que resultan imprescindibles para la regulación del crecimiento celular y la homeostasis. Entre otros, los receptores extracelulares que se encuentran habitualmente en esta vía están los receptores RTK, los receptores acoplados a proteínas G y receptores de citoquina (Samatar, 2014; Crespo Baraja, 2009; Rubinfeld, 2005; Dirk Bokemeyer, 1996).

Existen diferentes factores extracelulares que inician la activación de las ERK, como pueden ser los factores de crecimiento, incluidos el factor de crecimiento EGF, el factor PDGF, el factor NGF y también en respuesta a la insulina (Cargnello, Marie & Roux, Philippe P, 2011).

Muchos de los estudios realizados de las vías de señalización celular y particularmente de la vía de RAS-ERK, han evidenciado que las mutaciones prevalentes oncogénicas interrumpen el funcionamiento normal de estas vías, provocando con ello la desregulación de la mitogénesis, la resistencia a los daños pro-apoptóticos y una mejora en la motilidad, favoreciendo con ello el desarrollo de diferentes tipos de cáncer entre otras enfermedades (Nan, 2015; Samatar, 2014).

El mapa de vías MAPK presenta una dificultad elevada de análisis, dado que existe un entramado complejo de interrelaciones a través algunos de los componentes que actúan como hubs o elementos de conexión entre las diferentes rutas. Esto pone en evidencia la complejidad relativa al entendimiento completo del funcionamiento de cada una de estas vías y de sus mecanismos para su especificidad (KEGG, Consultado en 2018).

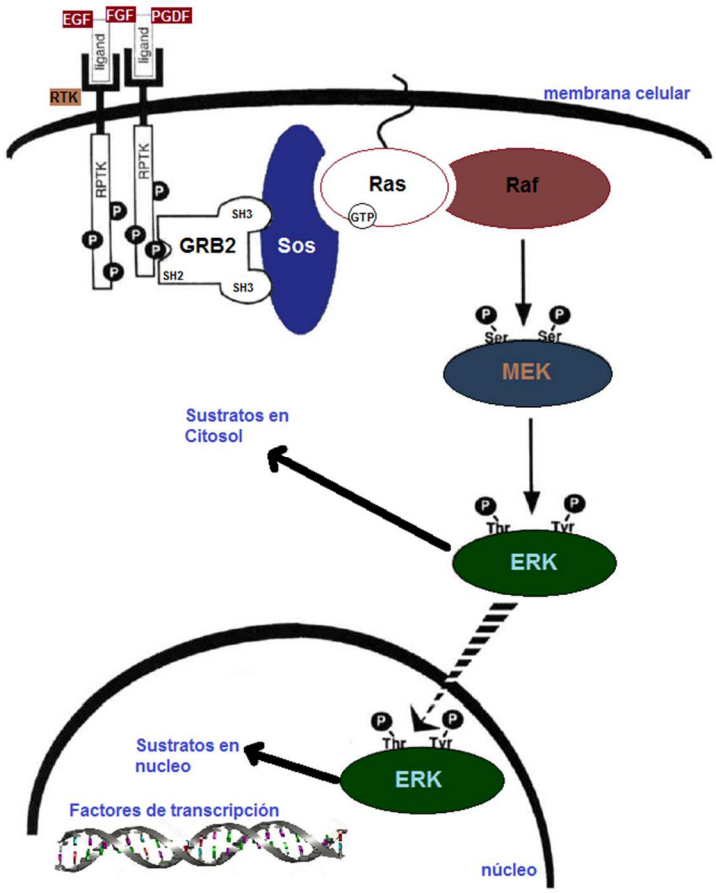


Figura F 2.3. *(Figura adaptada de Dirk Bokemeyer, 1996) Esquema de componentes de la ruta RAS-ERK*

2.2.1.1 Componentes y funcionamiento de la ruta

En la figura F 2.3 se muestran los principales componentes que intervienen en la activación de la ruta RAS-ERK. En la superficie celular se localizan los RTK, que es dimerizado inducido por el ligando (un factor de crecimiento) induciendo la activación del receptor y la autofosforilación de residuos de Tyr en el dominio intracelular. Los residuos fosforilados funcionan como sitios de unión específica para proteínas de dominios de homología Src 2 – SH2 (en inglés Src homology 2) o de unión a fosfotirosina (PTB), como la proteína 2 unida al receptor del factor de crecimiento, Grb2 (en inglés grown factor receptor-bound protein 2). Un factor de intercambio de nucleótidos de guanina (GEF), como es SOS (en inglés son of sevenless) media la activación de Ras, que se transloca desde el citosol hacia la membrana plasmática como resultado de su interacción con Grb2, estimulando así el intercambio de GDP unida a Ras por GTP, que es el que produce la regulación positiva de la actividad de Ras, es decir, activa Ras (ver figura F 2-4). (Dirk Bokemeyer, 1996; Karnoub, 2008).

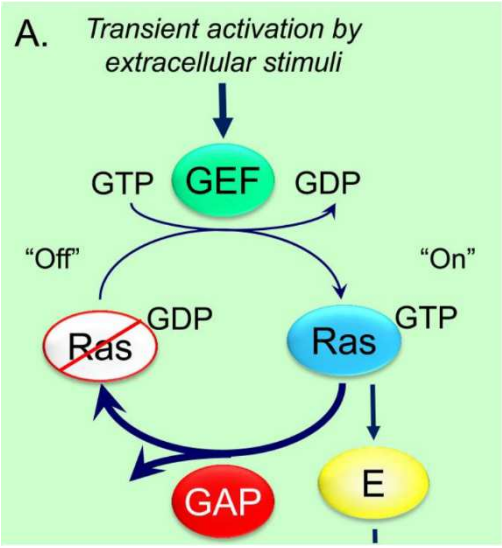


Figura F 2.4. *(Figura 2A de Baines, 2011) Ciclo de activación de Ras normal*

Para cerrar el ciclo de activación de Ras, el Ras activado es eventualmente inactivado por hidrolisis del GTP vinculado. El proceso de cambio de GDP y GTP es muy lento en ambas direcciones bajo condiciones fisiológicas sin catálisis por GAPs y GEFs, por lo cual se acelera con GAP (en inglés GTPase-accelerating proteins), el balance entre GAP y las actividades de GEF es un mecanismo de regulación crucial para el estado de activación de Ras (Takashima & Faller, 2013).

El intercambio de nucleótidos que se produce en este proceso, permite a Ras interactuar directamente con sus efectores diana, que son proteínas que tienen una gran afinidad con GTP-Ras. De los 10 efectores registrados, los dos más extensamente estudiados son Raf y PI_3K, aunque en este trabajo solo se profundizará en Raf (Takashima & Faller, 2013).

La familia de Raf comprende tres isoformas: A-Raf, B-Raf y C-Raf/Raf-1 (ver figura F 2.2). B-Raf es la MEK quinasa más fuerte y A-Raf es el activador de MEK más débil. Raf es la quinasa iniciadora del módulo ERK1/2. A-Raf activa preferentemente MEK1, mientras que B-Raf y Raf-1 activan ambos MEK1/2 con igual eficacia (Takashima & Faller, 2013). En el caso de Raf-1, por ejemplo, el Ras activado se une a la porción NH_2-terminal de la proteína serina-treonina quinasa Raf-1, reclutando este a la membrana plasmática. Raf-1 activa las isoformas de MAP quinasa quinasa MEK1 y MEK2 mediante la fosforilación de dos residuos reguladores de serina, que a su vez, fosforilan ERK1 (p44 MAP quinasa) y ERK2 (p42 MAP quinasa) en los sitios reguladores de treonina y tirosina, que están separados por un residuo de glutamato tanto en ERK1, como en ERK2 que dan el motivo TEY en el circuito de activación (Dirk Bokemeyer, 1996; Karnoub, 2008; Pearson, 2001).

Hay estudios que indican que el gen BRAF está mutado en el 66% de los melanomas y en el 12% de los cánceres colorrectales. Según el estudio realizado en el capítulo 3 de las muestras obtenidas en COSMIC v84, BRAF aparece mutado el 21.18% en tejidos de intestino grueso y el 46.09% en tejidos de tiroides. En cambio, es raro encontrar las mutaciones de Raf-1, A-Raf o MEK1/2 algún tipo de cáncer (Dhillon AS, 2007), según los datos analizados en el capítulo 3, solo aparecen en el 0.2% o menos de los tejidos analizados. A pesar de ello,

se ha verificado que las anomalías en esta vía conducen a la elevación del phospho-ERK1/2 (activado), se observa en numerosos cánceres humanos. Estos ERK activados se translocan al núcleo y activan factores de transcripción cuyos genes diana incluyen reguladores de la proliferación celular o regulación del ciclo celular, o, en algunos casos, reguladores de retroalimentación negativa de la vía Raf-MEK-ERK (Takashima & Faller, 2013).

2.2.2 Familia RAS y su implicación oncogénica

Como se ha indicado anteriormente, diferentes estudios de la vía RAS/ERK han evidenciado que mutaciones génicas en los diferentes componentes de la vía están presentes en un porcentaje importante en algunos tipos de cáncer. La implicación de la familia de proteínas RAS, como componente principal de esta vía en la carcinogénesis, está demostrada en innumerables estudios desde hace bastantes años, por ejemplo, en el 35% de adenocarcinoma de pulmón (células no pequeñas), en el 90% de cáncer de páncreas, o en el 45% de cáncer colorectal, entre otros (Cargnello, Marie & Roux, Philippe P, 2011).

Las proteínas de la familia RAS tienen funciones esenciales en las vías de señalización celular, que regulan el crecimiento celular. Se ha comprobado que la activación de la proteína RAS, debido a mutaciones activadoras del gen que la codifica, contribuye de forma significativa al fenotipo maligno de las células, a causa de la desregulación en las funciones esenciales de supervivencia, proliferación y diferenciación de las células tumorales, y con ello alterando las capacidades de la muerte celular programada, el potencial de invasión y la capacidad de inducir nuevos vasos sanguíneos (Crespo Baraja, 2009; Cargnello, Marie & Roux, Philippe P, 2011).

2.2.2.1 Activación crónica de RAS

La activación y desactivación de la proteína RAS son dos procesos esenciales en la vía de señalización RAS-ERK, dado que con un funcionamiento alterado promueven la activación de RAS de forma crónica y por tanto toda la cascada de señalización aguas abajo. El ciclo normal de activación de RAS, figura F 2.4, muestra como las proteínas RAS de tipo salvaje van pasando de un estado de unión a GTP activo, a un estado de unión GDP inactivo, es decir, actúan como un interruptor binario (encendido/apagado) que regulan la señal a través de la cascada de señalización. Los factores de intercambio de guanina GEF (Guanine Exchange Factor) promueven la formación de las proteínas RAS-GTP activadoras, por el contrario, las proteínas GTPasa (GAP) promueven la formación del RAS-GDP inactivo. Las mutaciones oncogénicas de RAS presentan sustituciones sin sentido de un único aminoácido, que se localiza comúnmente en los residuos 12, 13 o 61, que las insensibilizan a GAP y que producen la estabilización de su unión a GTP, provocando así la activación crónica de RAS y por tanto, de las cascadas de señalización aguas abajo. Ver figura F 2.5 (Samatar, 2014; Baines, Xu, Dapeng, & Der, Channing J, 2011).

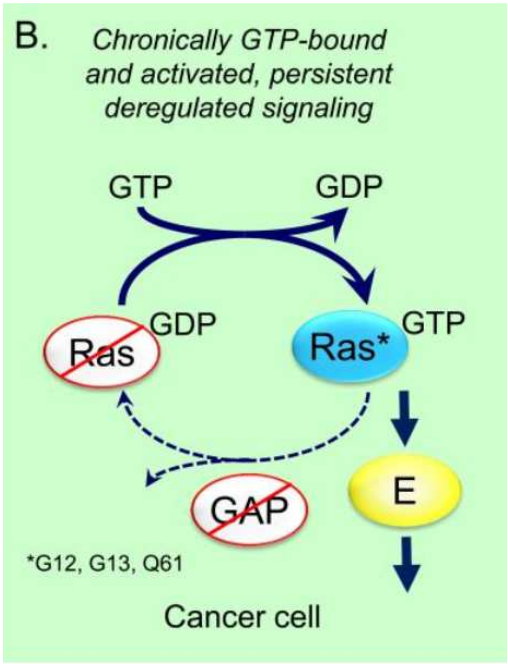


Figura F 2.5. *(Figura 2B de Baines, 2011) Activación crónica de Ras mutado*

En un principio, se pensaba que RAS se ubicaba únicamente en la parte interna de la membrana plasmática, pero estudios más recientes han mostrado que la señalización de RAS se ha encontrado en membranas intracelulares como los endosomas, el retículo endoplásmico, el aparato de Golgi y las mitocondrias. Por ello, el RAS activado se puede unir y estimular un espectro heterogéneo de efectores funcionalmente diversos aguas abajo, dando con ello la activación regulada de una matriz compleja de redes de señalización citoplásmica (Baines, Xu, Dapeng, & Der, Channing J, 2011).

En base al papel relevante de la familia de las proteínas RAS en los procesos de oncogénesis de muchos tipos de cáncer, RAS se ha convertido en uno de los objetivos principales de estudio para conseguir su bloqueo y evitar la cronicidad de su activación. En el siguiente apartado se profundiza en el análisis de la familia de proteínas RAS, en sus diferentes isoformas, su ubicación en el mapa de señalización RAS y en los diferentes tipos de mutaciones encontradas en ellas.

2.2.3 Estructuras del dominio de las proteínas RAS

En HRAS, NRAS y KRAS los dominios entre los aminoácidos 1 al 165, dominios catalíticos amino-terminales, están muy conservados, por encima del 90% de identidad. Entre los aminoácidos 166 a 188/189 están las secuencias carboxi-terminales, que divergen significativamente entre cada proteína y que se conocen como domino hipervariable (HVR). El domino HVR (figura F 2.6) está definido por una secuencia proteica llamada enlazador (*linker* en inglés), situada entre el dominio conservado y la secuencia de anclaje. Las secuencias de anclaje están bien caracterizadas, donde el anclaje RAS mínimo está formado por el motivo CAAX carboxi-terminal y una segunda señal. La segunda señal comprende un único sitio de palmitoilación (C181) en NRAS, dos sitios de palmitoilación en HRAS (C181 y C184) y un dominio polibásico de seis residuos de lisina contiguos en KRAS (K175 – K180) (Hancock, 2003).

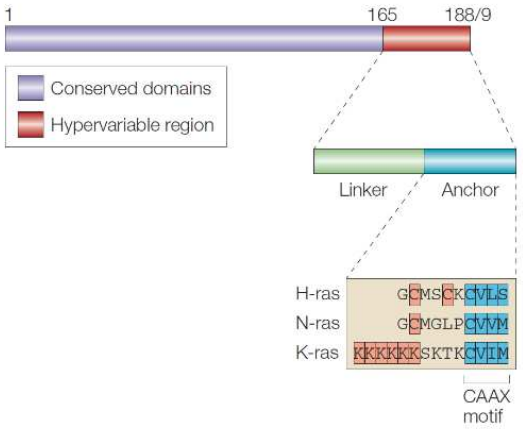


Figura F 2.6. *(Figura 1 de Hancock, 2003) Estructura de dominio de RAS*

Las cuatro proteínas codificadas por los genes RAS (teniendo en cuenta las 2 isoformas de KRAS) normalmente se encuentran unidas a la lámina interna de la membrana plasmática, con la finalidad de que puedan realizar plenamente sus funciones. Esta unión se realiza mediante adiciones lipídicas post-traduccionales en el extremo carboxilo terminal del motivo CAAX de HVR (Crespo Baraja, 2009).

2.3 Mutaciones en KRAS: subtipos según mutación

2.3.1 Oncogenes RAS humano: HRAS, NRAS y KRAS.

La familia RAS es un subconjunto de *RAS small GTPase superfamily*. Actualmente, la familia RAS contiene 31 genes, entre los cuales incluye los protooncogenes HRAS, NRAS y KRAS. En la figura F 2.7 se muestra el mapa jerárquico de las familias de genes, donde se encuentra la familia de gens RAS (HGNC, 2018).

Los genes RAS que se descubrieron en primer lugar fueron HRAS y KRAS, que se identificaron a partir de estudios sobre el cáncer que realizaron Edward M. Scolnick y su equipo en el NIH (National Institutes of Health), de dos virus causantes de cáncer, el virus del sarcoma Harvey y el sarcoma de Kirsten, por Edward M. Scolnick. Posteriormente se descubrieron los genes RAS humanos activados y transformados en células cancerosas humanas en Harvard. Un tercer gen RAS, llamado NRAS, fue descubierto posteriormente por investigadores del Institute of Cancer Research, su nombre hacía referencia inicialmente a la identificación de las células del neuroblastoma humano.

Existe una homología muy alta entre los tres genes RAS humanos, entre el 85% y el 90% (Downward, 2003; Baines, Xu, Dapeng, & Der, Channing J, 2011) formadas por cadenas de 188 a 189 aminoácidos. Sus símbolos genéticos son HRAS, NRAS y KRAS, el último de los cuales produce las isoformas K-Ras4A y K-Ras4B de splicing alternativo (HGNC, 2018).

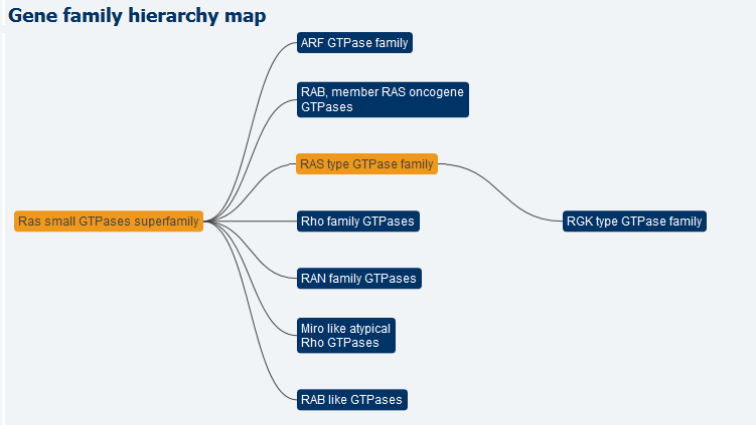


Figura F 2.7. *(HGNC, 2018) Mapa de jerárquico de la familia de genes RAS*

HRAS

La localización cromosómica de HRAS es 11p15.5 y las coordenadas en el genoma humano (GRCh38/hg38) son chr11:532,243-535,550. Los defectos en este gen están implicados en una variedad de cánceres, incluidos el cáncer de vejiga, el cáncer folicular de tiroides y el carcinoma oral de células escamosas. En la figura F 2.8 se puede ver el gen HRAS en el navegador genómico UCSC, donde se muestran gráficamente la localización del gen en el cromosoma 11 humano (ensamblaje GRCh38/hg38) (UCSC, 2018; HGNC, 2018).

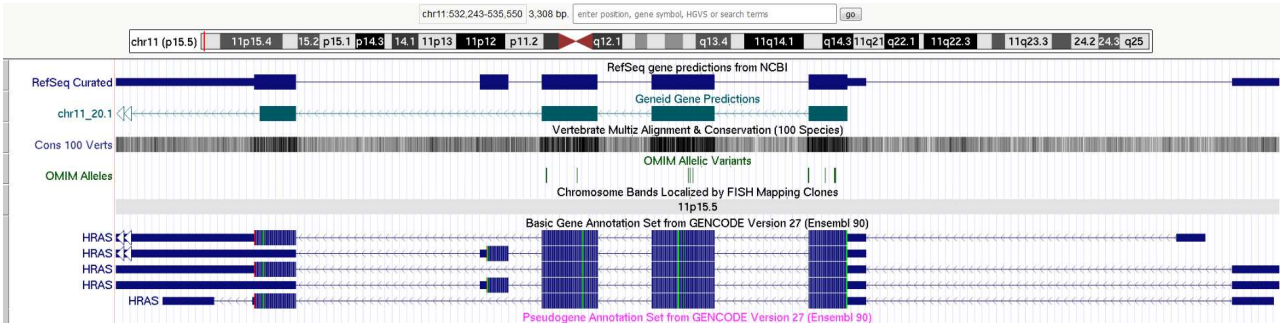


Figura F 2.8. *(UCSC, 2018) Gen HRAS en el navegador genómico de UCSC*

KRAS

El gen KRAS está localizado cromosómicamente en 12p12.1 y sus coordinadas en genoma humano (GRCh38/hg38) son chr12:25,209,431-25,250,803. En este gen, una única sustitución en un aminoácido, genera una mutación donde la proteína resultante (mutada) está implicada en diversas enfermedades malignas, que incluyen adenocarcinoma de pulmón, adenoma mucinoso, carcinoma ductal de páncreas y carcinoma colorrectal. El corte y empalme alternativo conduce a variantes que codifican dos isoformas que difieren en la región C-terminal, las K-Ras4A y K-Ras4B. En la figura F 2.9 se puede ver el gen KRAS en el navegador genómico UCSC, donde se muestran gráficamente la localización del gen en el cromosoma 12 humano (ensamblaje GRCh38/hg38) (UCSC, 2018; HGNC, 2018) .

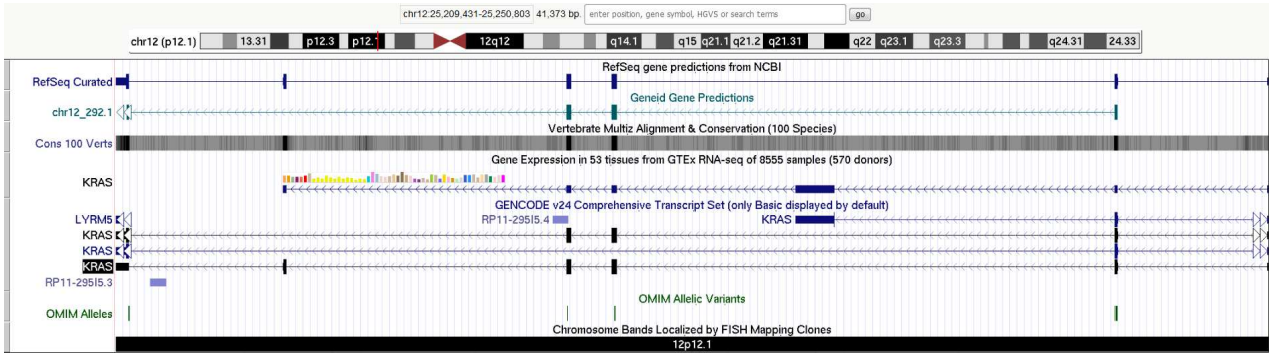


Figura F 2.9. *(UCSC, 2018) Gen KRAS en el navegador genómico de UCSC*

NRAS

Este es un oncogen NRAS que codifica una proteína de membrana que se ubica entre el aparato de Golgi y la membrana plasmática. El oncogen NRAS está localizado cromosómicamente en 1p13.2 y sus coordinadas en genoma humano (GRCh38/hg38) son chr1:114,704,469-114,716,894. Las mutaciones en este gen se han asociado con el cáncer de recto somático, el cáncer de tiroides folicular, el síndrome linfoproliferativo autoinmune, el síndrome de Noonan y la leucemia mielomonocítica juvenil. En la figura F 2.10 se puede ver el gen KRAS en el navegador genómico UCSC, donde se muestran gráficamente la localización del gen en el cromosoma 1 humano (ensamblaje GRCh38/hg38) (UCSC, 2018; HGNC, 2018).

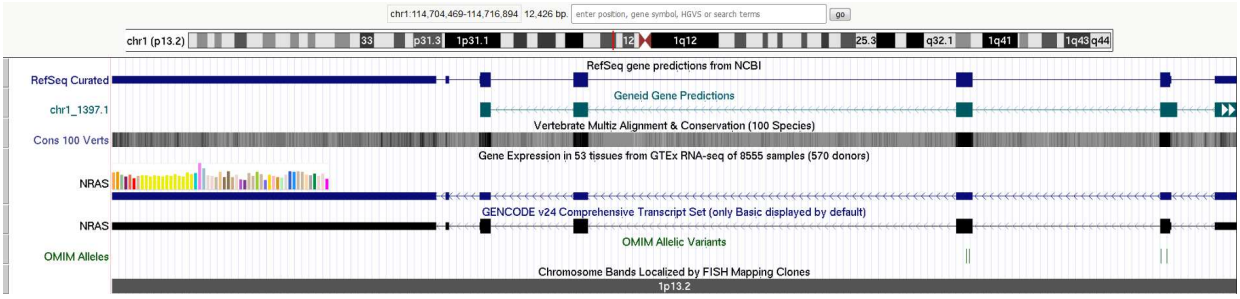


Figura F 2.10. *(UCSC, 2018) Gen NRAS en el navegador genómico de UCSC*

2.3.2 Mutaciones en RAS.

El funcionamiento anormal de las vías de señalización de RAS se produce entre otras por la mutación de los genes RAS. Sobre el 33% de los cánceres humanos tienen mutaciones puntuales que afectan a la activación en RAS. KRAS tiene el 86% de todas las mutaciones de RAS y con el 21.6% KRAS es el de mayor frecuencia de todos los cánceres humanos, seguido de NRAS con aproximadamente el 8 % y en último lugar HRAS con menos de un 3.3% (Baines, Xu, Dapeng, & Der, Channing J, 2011). En los datos obtenidos en el análisis realizado en el capítulo 3 de muestras de tejido con diferentes tipos de cáncer (COSMIC v84, 2018), ver gráfica Gr. 3.5, los porcentajes resultantes de mutación han sido del 31.9% de KRAS, del 4.8% de NRAS y del 1.4% de HRAS.

Las mutaciones oncogénicas de RAS son sustituciones de bases únicas, que principalmente afectan a los residuos G12 (Gly), G13 (Gly) y Q61 (Gln), aunque no en la misma proporción para cada gen HRAS, NRAS o KRAS. Principalmente

estas mutaciones conducen a la estabilización de la unión RAS-GTP y la activación crónica de RAS y cascadas de señalización aguas abajo (Crespo Baraja, 2009).

2.3.2.1 Distribución de mutaciones para RAS.

En base a los datos recogidos de muestras con mutaciones de los genes RAS (HRAS, NRAS y KRAS) en COSMIC v84 (Cataloge of Somatic Mutations In Cancer, consulta realizada el 1 de abril d 2018), el 99.1% de las mutaciones son por sustituciones de una sola base, de las cuales las más frecuentes son las producidas en los codones 12, 13 y 61, pero con distribuciones diferentes para cada una de las tres proteínas.

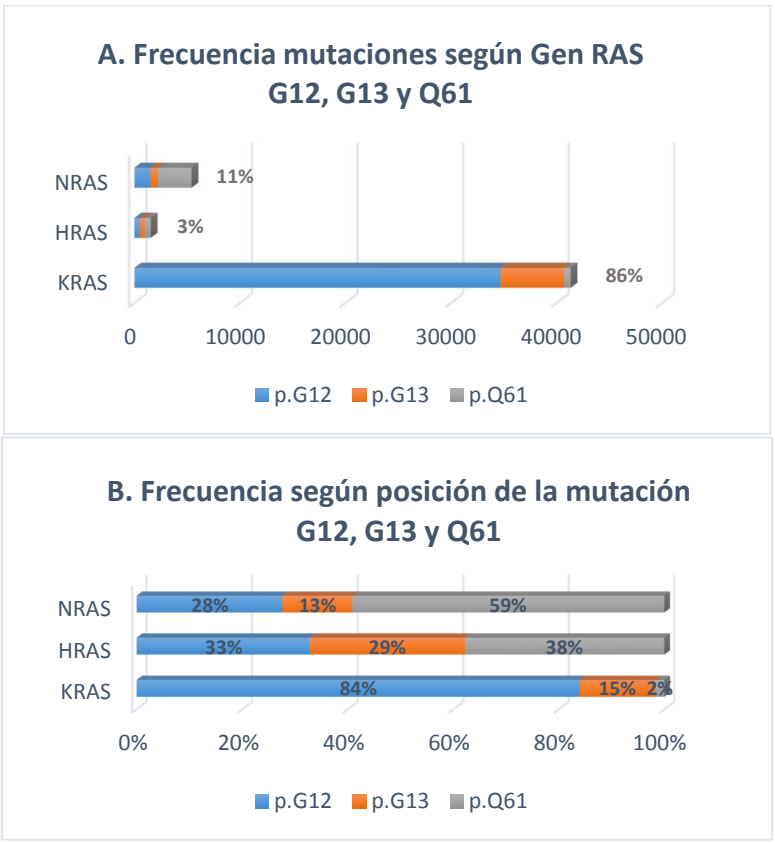


Figura F 2.11. *(COSMIC v84, 2018) Distribución de mutaciones por sustitución en codones 12, 13 y 61. A: Distribución de frecuencias por Gen. B: Distribución de frecuencias por posición de la mutación en cada gen*

En las gráficas de la figura F 2.11 se puede ver esta distribución. Como se puede apreciar en la gráfica A, las mutaciones en KRAS son las más frecuentes, seguida de NRAS y en último lugar HRAS. Adicionalmente en la gráfica B, se puede comprobar que las mutaciones más abundantes en KRAS son en los codones 12, en primer lugar y 13, en segundo lugar. En NRAS esta distribución varía y la frecuencia más alta de aparición de la mutación es en el codón 61, en primer lugar y en el 12, en segundo lugar. En HRAS la distribución es por igual en los tres codones, aunque el volumen de apariciones de HRAS mutado es muy inferior a KRAS y NRAS (COSMIC v84, 2018).

El cáncer de pulmón, de colon y de páncreas son tres de los cuatro principales neoplasmas que causan la muerte de cáncer en EEUU, el 25.9%, 8.4% y 7.2% respectivamente, que presentan el gen predominante mutado KRAS (Siegel, Miller, & Jemal, 2017). Analizando la base de datos de COSMIC v84 (Cataloge of Somatic Mutations In Cancer), se puede comprobar que de 39,083 muestras el cáncer de pulmón, donde se ha testeado la mutación de gen KRAS, 6,843 muestras han dado positivo en alguna de las mutaciones de KRAS, es decir, un 17.5% de las muestras de este tipo de cáncer tiene mutación de KRAS (COSMIC v84, 2018). Más adelante, en el Capítulo 3 de este trabajo se profundiza en las estadísticas sobre la incidencia del cáncer de pulmón respecto a otros, en referencia a la mutación de KRAS.

Como se ha podido comprobar, las mutaciones en KRAS son las más frecuentes en algunos de los cánceres con más mortalidad, incluido el cáncer de pulmón, por lo que este trabajo se centra en este ámbito concreto: mutación en KRAS y su implicación en el cáncer de pulmón. En los siguientes apartados se profundiza en esta línea.

2.3.2.2 Tipos de mutaciones en KRAS.

Tal como se ha indicado anteriormente, las mutaciones más frecuentes en KRAS se producen principalmente en el residuo G12 y en menor frecuencia en G13. En la gráfica F 2.12 se muestra la distribución de frecuencia según el tipo de mutación producido en cada uno de esos dos codones. Las mutaciones más frecuentes en el codón 12 son, G12D (cambio a Asp) con un 42%, seguida de la G12V (cambio a Val) con un 28% y la tercera la G12C (cambio a Cys) con un 14%. En el codón 13 solo destaca la G13D (cambio a Asp) con una frecuencia del 89% respecto a los otros tipos de mutación (COSMIC v84, 2018).

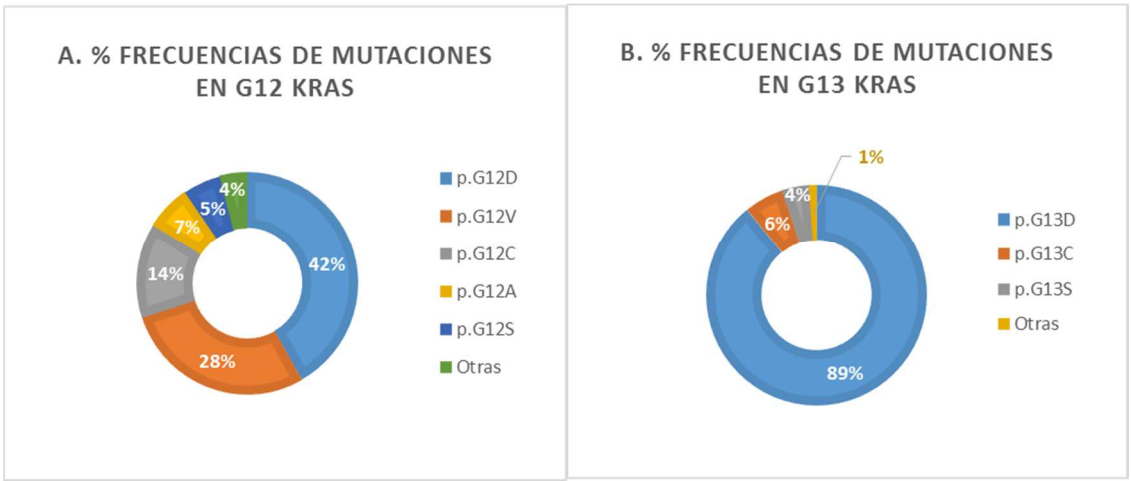


Figura F 2.12. *(COSMIC v84, datos de consulta realizad 1 de abril de 2018) Distribución de tipos de sustituciones en residuos G12 y G13. A: Distribución de frecuencias G12. B: Distribución de frecuencias en G13.*

Las sustituciones de las bases en las secuencia de cDNA de las mutaciones en los codones 12 y 13, se pueden ver en la gráfica F 2.13, donde se observan las distintas sustituciones que se producen en cada base nitrogenada y que provocan el cambio de aminoácidos según las frecuencias arriba descritas.

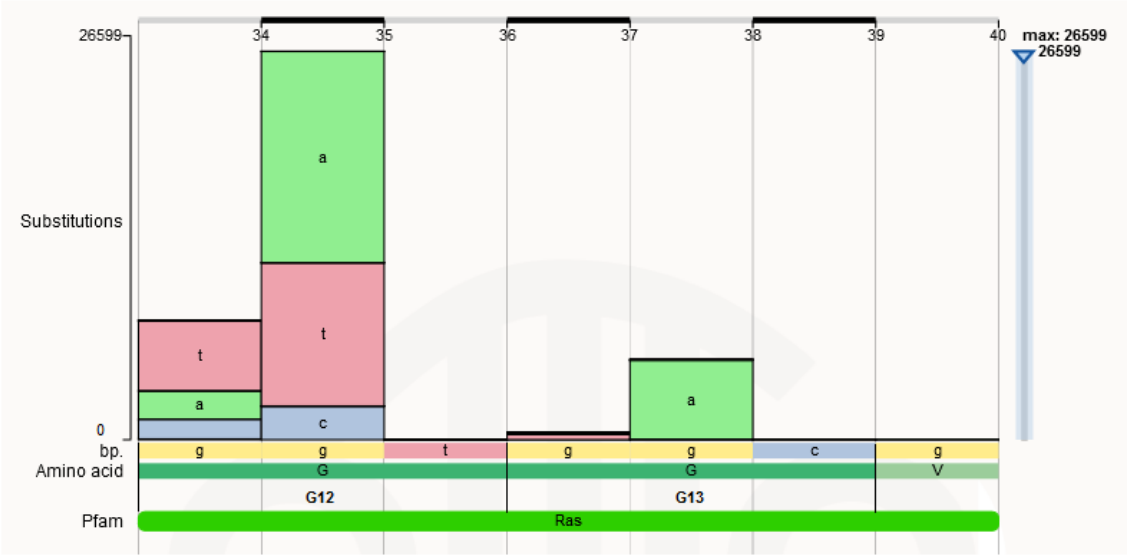


Figura F 2.13. *(Grafico adaptado de COSMIC v84, consulta realizada 1 de abril de 2018) Sustituciones en G12 y G13 de KRAS.*

Las mutaciones de KRAS relacionadas con el cáncer se producen en el codón 12 y en el codón 13, como se ha comentado anteriormente. Para profundizar sobre la relación entre el tipo de mutación existente en KRAS y el tipo de cáncer sobre el que incide, se ha realizado un estudio de la base de datos de mutaciones somáticas en el cáncer COSMIC V84 (consulta realizada el 2 de abril de 2018), cuyo resultado se muestra en las tablas 2.1 A y B. Se han seleccionado solo los cánceres que tienen una incidencia superior a 100 casos en la mutación somática de KRAS y respecto a la mutación, solo se han tenido en consideración las ubicaciones en los codones 12 y 13. Por otra parte, también se han desestimado los tipos de mutación dentro de cada ubicación que tenían una incidencia menor a 10 casos por tipo.

Se puede destacar que los tres primeros tipos de cáncer donde la mutación de KRAS parece tener una incidencia destacada respecto a los otros, es en los tejidos de intestino grueso, pulmón y páncreas, en ese orden. En intestino grueso las mutaciones más prevalentes son la G12D y G12V, al igual que en páncreas, pero en el intestino grueso también parece destacar la mutación en el codón 13, G13D, mientras que en pulmón y páncreas no. En pulmón, a diferencia de intestino grueso y páncreas, destaca la mutación G12C, aunque también aparece con un cierto valor la G12V. En este caso no parecen ser destacadas las mutaciones en el codón 13.

***Tabla 2.1**. Mutaciones somáticas de KRAS en codón 12 (A) y 13 (B) en cáncer.*
Con fondo crema se indican los valores más significativos en cada tipo de tejido. (Datos recogidos de COSMIC v84, 2018)

A	**Codón 12**								
Tejido primario	**G12A**	**G12C**	**G12D**	**G12F**	**G12G**	**G12L**	**G12R**	**G12S**	**G12V**
Large intestine	1.415	1.932	8.411	20	4	1	298	1.363	5.412
Lung	504	2.342	1.277	28	2	2	103	205	1.488
Pancreas	84	174	3.034	2	7	8	802	110	1.971
Ovary	45	44	321				30	11	310
Biliary tract	34	53	370				27	69	123
Haematopoietic and lymphoid	53	26	218		2		22	48	74
Endometrium	68	54	219				4	17	125
Stomach	27	16	107	1			1	12	38
Small intestine	16	16	97				3	16	36
Peritoneum	1	11	95				1	2	36
Thyroid	6	18	39				19	20	20
Prostate	2	10	22				1	3	40
Soft tissue	5	17	29	1				1	6
Total	**2.260**	**4.713**	**14.239**	**52**	**15**	**11**	**1.311**	**1.877**	**9.679**

B	**Codón 13**							
Tejido primario	**G13A**	**G13C**	**G13D**	**G13G**	**G13R**	**G13S**	**G13V**	**V14I**
Large intestine	19	119	4.500	10	47	29	28	13
Lung	3	184	188		6	6	5	1
Pancreas	3	5	105	2	4	7	2	2
Ovary	4	7	37		1	1	3	
Biliary tract		6	31	1	4	3		
Haematopoietic and lymphoid	2	8	211	4	1	1		3
Endometrium	2	9	72	2		1	2	1
Stomach		5	58	1		11	1	1
Small intestine			36				1	1
Peritoneum		2	16		1			1
Thyroid	3	1	24		2	11		
Prostate			23			3		
Soft tissue		1	36		1	3	1	
Total	**36**	**347**	**5.337**	**20**	**67**	**76**	**43**	**23**

Según los valores mostrados en la tabla 2.11, las mutaciones somáticas en KRAS en el cáncer de pulmón se produce principalmente en el codón 12 (94%), en tres tipos de forma más habitual: G12C con un 37%, G12D con un 20% y G12V con un 23%.

2.3.2.3 Etiología de las mutaciones de KRAS.

En referencia a la etiología de las mutaciones de KRAS, hay estudios de carcinogénesis química que indicaron que la metilnitrosourea (MNU) se dirige a la segunda base del codón 12 de HRAS y KRAS para generar mutaciones del tipo G12D en diferentes tipos de cáncer (Prior & Lewis, 2012). La mutación G12C parece estar fuertemente correlacionada con el cáncer de pulmón, según se ha podido ver en COSMIC v84 (consulta realizada 1 de abril de 2018), el 34.5% de los cánceres de pulmón registrados a la fecha de la consulta, presentan la mutación G12C en KRAS. Este tipo de mutación G12C se han asociado en estudios *in vitro* con la formación de aductos voluminosos de ADN generados por productos del humo del tabaco (Seo, Jelinsky, & Loechler, 2000). Por otra parte, también se relaciona una alta frecuencia de las mutaciones en el codón 13 de KRAS y NRAS con el cáncer colorrectal (CCR) y los cánceres hematopoyéticos / linfoides (Prior & Lewis, 2012).

2.3.2.4 Mutaciones de KRAS en el cáncer de pulmón.

En Europa, el cáncer de pulmón representa el 12,1 % (449 K) de nuevos casos de cáncer y el 20,1 % (388 K) de muertes relacionadas con el cáncer. (IARC - World Health Organitation, 2012; NIH - National Cáncer Institute, 2018). La proporción de muertes por cáncer de pulmón versus a los nuevos casos diagnosticados, es muy alta, en Europa representa el 86,4 % aproximadamente. En el cáncer de pulmón, KRAS es uno de los genes más mutados de la familia RAS, aunque también en algunos pocos casos HRAS y NRAS aparece mutado.

Según un estudio realizado mediante análisis de regresión logística para establecer relación entre variables clínicas y patológicas con las mutaciones de KRAS, mostró que las mutaciones de KRAS eran significativamente más comunes en pacientes fumadores con adenocarcinomas primarios de pulmón, que no en pacientes no fumadores. En base al sexo de los pacientes, las proporciones en las mutaciones KRAS eran similares. No se encontraron diferencias significativas ni por sexo (hombres vs mujeres), ni por edad, ni por el tipo de histología celular, el estadio patológico o el tamaño del tumor (Steven A. Ahrendt et al., 2001).

Realizando un análisis de muestras de cáncer de pulmón con mutación de KRAS positiva (6,843 muestras), la histología mayoritaria que aparece es el *Carcinoma* con un 97.3% de incidencia, del cual se distinguen dos subtipos más frecuentes: el *Adenocarcinoma de pulmón*, con un 51.8% de incidencia y el *Carcinoma de pulmón célula no pequeña CPCNP (NSCLC - non-small-cell lung cáncer* en inglés*)*. Las mutaciones KRAS más habituales en los dos subtipos, son por orden de frecuencia G12C, G12V y G12D. En la figura F 2.14 se pueden ver las proporciones en ambos casos (datos recogidos de COSMIC v84, 2018).

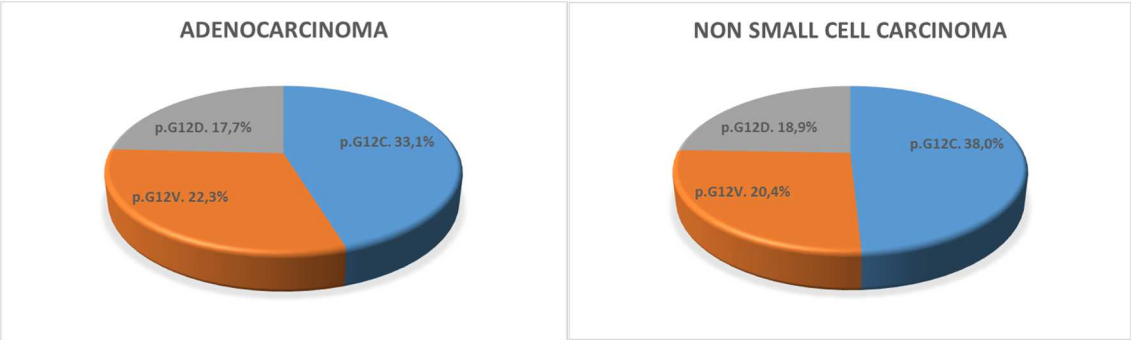


Figura F 2.14. *Mutaciones KRAS en Adenocarcinoma de pulmón y en NSCLC*

Según la revisión estudios recientes realizado por A. Fernández y su equipo (McCormick, 2015), se sugiere que podría existir una relación entre el pronóstico final y el tipo de mutación KRAS. Por ejemplo, se menciona que las mutaciones G12D se asocian con tumores con mejor pronóstico que las que las que presentan G12V o G12R, aunque la frecuencia de esta última mutación es muy baja. También parece existir una asociación de la presencia de mutaciones en KRAS con una falta de respuesta a los inhibidores de EGFR para el CPCNP (David A. Eberhard et al., 2005; Pao, William et al., 2005). Hay estudios realizados sobre ratones, que han demostrado que los ratones que carecían de la isoforma KRAS4A eran altamente resistentes al desarrollo de tumores pulmonares inducidos por carcinógenos. Esto sugiere que KRAS4A puede ser esencial para la carcinogénesis pulmonar (To, 2008).

2.4 Dianas terapéuticas relacionadas con KRAS

Las mutaciones de RAS en los exones 12, 13 y 61 afectan a la activación de la vía de señalización RAS-ERK aguas abajo, de forma que estabilizan la unión de la proteína RAS a GTP, generando la cronicidad de la activación de las vías. Como se ha podido ver, tanto en el mapa de las vías de MAPK (figura F 2.2), son múltiples los efectores de RAS funcionalmente diversos que intervienen aguas abajo, pero aún no se conoce con certeza si los efectos oncogénicos de RAS son solo generados por uno de estos efectores o por la acción colectiva de todos ellos a través de las múltiples vías interconectadas (Samatar, 2014; Prior & Lewis, 2012). Dada esta complejidad, se han ido estableciendo diversas estrategias para intentar bloquear el proceso de oncogénesis mediado por RAS.

Varios son los enfoques que a lo largo de los últimos 30 años se han ido aplicando al proceso de búsquedas de dianas terapéuticas dirigidas al oncogén RAS desde el descubrimiento de la existencia de mutaciones en la familia de proteínas RAS (KRAS, NRAS y HRAS). En muchos de los estudios realizados en estos años, se pueden identificar varios enfoques: i) apuntar a RAS directamente (o a los pasos de su activación), ii) apuntar a la ruta RAS aguas abajo (Samatar, 2014; Crespo Baraja, 2009; Takashima & Faller, 2013; Prior & Lewis, 2012); iii) apuntar a la *letalidad sintética ("synthetic lethality")* con el objetivo de atacar (en lo posible) directamente a las células cancerígenas, sin alterar o dañar las células normales (Takashima & Faller, 2013; Kaelin, 2009), iv) el estudio de genes/proteínas relacionadas con la inhibición o activación de las vías de señalización impulsadas por KRAS como dianas terapéuticas, para

evitar la retroalimentación en la señalización celular. Seguidamente, se describirán brevemente las diferentes estrategias utilizadas con esta finalidad, especialmente el enfoque iv), que será la base para el análisis bioestadístico del capítulo 4 de la expresión de diferentes genes y su influencia en la supervivencia de los pacientes. (Samatar, 2014; Crespo Baraja, 2009; Takashima & Faller, 2013; Baines, Xu, Dapeng, & Der, Channing J, 2011; Gil Moreno, 2018).

2.4.1 Diana: vías de activación de RAS.

2.4.1.1 Inhibidores de asociación de RAS a la membrana plasmática.

Partiendo del hecho de que para que se produzca la activación de RAS es imprescindible su unión a la membrana plasmática, uno de los primeros enfoques aplicados para inhibición de la activación consecutiva, en el cual se depositaron muchas expectativas, se basó en dirigir los esfuerzos de investigación en bloquear esta unión. En este sentido, se han estudiado diferentes inhibidores que consiguieran este efecto: inhibidores de farnesiltrasferasa (FTI), inhibidores de geranil geraniltransferasa (GGTI), inhibidores de moléculas pequeñas e inhibidores de fosfodiesterasa d (PDEd).

2.4.1.1.1 Inhibidores de FTI, de GGTI y de PDEd

Aunque inicialmente se dedicaron muchos esfuerzos de investigación a este tipo de inhibidores, posteriormente se vio que los inhibidores de FTI no bloqueaban eficazmente la prenilación de las proteínas KRAS y NRAS para evitar su activación (Baines, Xu, Dapeng, & Der, Channing J, 2011). *Tipifarnib* y *lonafarnib* son dos de los inhibidores FTI que han avanzado a ensayos de Fase III, pero donde hasta el momento no se han obtenido buenos resultados para los cánceres evaluados, entre ellos el CPCNP (Takashima & Faller, 2013). En base a estas experiencias, se determina que los inhibidores FTI no son eficientes para mutaciones en KRAS y llevó al intento de combinar FTI con inhibidores de geranil gerniltransferasa GGTasa (GGTI), para inhibir la activación de KRAS, pero que según parece, pueden producir toxicidad no deseada para las células normales (Takashima & Faller, 2013). Más recientemente, se ha propuesto una nueva diana, la proteína de unión a prenil PDEd, para alterar la localización subcelular y la activación de RAS. Al parecer suprimen la activación de KRAS oncogénica en modelos de laboratorio de cáncer inducido por KRAS (Zimmermann, 2013).

2.4.1.1.2 Otros procesos de inhibición de la asociación de RAS a la membrana

Algunas otras estrategias que se han intentado para evitar que RAS permanezca unido a la membrana plasmática, han estado orientadas a identificar inhibidores de Rce1 e inhibidores de isoprenilcisteína carboximetiltransferasa (ICMT) (Prior & Lewis, 2012).

Otras moléculas pequeñas que contienen grupos de Farnesyl (*salirasib* y *TLN-4601*) se han evaluado en ensayos clínicos como posibles inhibidores de la asociación de la membrana Ras y la oncogénesis (Baines, Xu, Dapeng, & Der, Channing J, 2011). Otro compuesto que está actualmente en estudio es el *ARS-853*, un inhibidor covalente selectivo de KRAS (G12C) que inhibe la señalización inducida por KRAS mutante (Patricelli, 2016). El tratamiento con *ARS-853*

disminuye en un 95% la unión entre GTP y KRAS, disminuyendo también la actividad de CRAF, AKT, ERK y aumentando la apoptosis (Gil Moreno, 2018).

2.4.2 Diana: vías efectoras de RAS aguas abajo.

2.4.2.1 Inhibidores de RAF.

Los inhibidores RAF surgen de la búsqueda de dianas terapéuticas que puedan bloquear la activación crónica de la ruta RAF-MEK-ERK. El primer inhibidor de RAF aprobado por la FDA fue sorafenib, que se diseñó como inhibidor para BRAF salvaje y CRAF, aunque en menor potencia también inhibe el mutante BRAF oncogénico (V600E) (Samatar, 2014).

Vemurafenib y *dabrafenib* fueron los primeros inhibidores de este tipo que entraron en ensayos clínicos. El resultado *in vivo* fue que inhiben la actividad de RAF y la señalización aguas abajo preferentemente en la forma mutante BRAF (V660E). En el caso de las células con BRAF de tipo salvaje, estos inhibidores de RAF, no inhiben RAF, sino que lo activan y también la señalización ERK (efecto conocido como *paradoja del inhibidor de RAF*) aguas abajo en varios grados, con efectos secundarios no deseables (Takashima & Faller, 2013; Samatar, 2014).

2.4.2.2 Inhibidores de MEK.

El mecanismo de acción de los inhibidores de MEK se basa en la capacidad que tienen de unirse a MEK y bloquearla en estado catalíticamente inactivo. Dado que las secuencias a las que se unen son únicas, son altamente específicas (Crespo Baraja, 2009). El primer inhibidor de MEK que entró en ensayos clínicos fue un inhibidor alostérico altamente específico, identificado como *CI-1040* (Crespo Baraja, 2009; Takashima & Faller, 2013; Samatar, 2014). A partir de la optimización de *CI-1040* resultó un inhibidor análogo identificado como *PD032590. S*e detuvo su desarrollo a causa de las toxicidades de la oclusión de la vena retiniana que producía (Samatar, 2014). *Selumetinib* (*AZD6244*) es otro inhibidor alostérico de MEK1 y MEK2, que mostró una actividad antitumoral en varios modelos de xenoinjertos humanos, incluyendo colon, páncreas, mama, NSCLC y melanoma, en la evaluación preclínica y se trasladó al desarrollo clínico (Baines, Xu, Dapeng, & Der, Channing J, 2011).

2.4.2.3 Inhibidores de ERK.

Inicialmente, el desarrollo de inhibidores de ERK fue limitado, por la suposición de que ERK era el único objetivo conocido de MEK aguas abajo en la vía y por tanto, este no tendría ningún beneficio adicional al inhibidor de MEK. Tras la experiencia con los inhibidores de RAF y MEK, su estudio se ha intensificado recientemente, dados los problemas de bucles de retroalimentación negativa y la frecuente recuperación de la señalización ERK debida a la resistencia de RAF y los inhibidores de MEK (Aronov, 2009; Hatzivassiliou, 2012; Ohori, 2005). Como ejemplo de este tipo de inhibidores está *SCH772984,* es un inhibidor competitivo de ATP de ERK que al unirse a ERK genera un doble mecanismo de

inhibición (Samatar, 2014; Nissan, 2013). Actualmente se están probando nuevos inhibidores ERK en ensayos clínicos de Fase I, *MK-8353 (SCH900353), BVD-523 y RG7842* (Samatar, 2014).

2.4.2.4 Combinación de Inhibidores.

Se están realizando ensayos clínicos que combinan inhibidores de RAF y MEK, partiendo de la observación de que la resistencia al inhibidor RAF (vemurafenib) con frecuencia está causada por la reactivación de la señalización de ERK. Al aplicar un inhibidor del dímero RAF conjuntamente con un inhibidor de MEK, se consigue un mayor efecto antitumoral en modelos preclínicos, al producir una supresión más efectiva y duradera de la señalización de ERK (Flaherty, 2012).

También se ha probado que los inhibidores de ERK en combinación con inhibidores de RAF son menos tóxicos que el uso individual de cada inhibidor, dado que los inhibidores de ERK y RAF tienen actividad opuesta en células normales (Nissan, 2013).

2.4.3 Diana: interacción letal sintética.

Aunque el concepto de “letalidad sintética” tiene más de 60 años, este concepto está resurgiendo como una estrategia emergente para el desarrollo de enfoques indirectos hacia las células mutantes cancerosas, por ejemplo en RAS, donde se ha visto que hay dificultades en abordar directamente las proteínas RAS como terapia contra el cáncer (Kaelin, 2009).

2.4.4 Diana: Genes relacionados con vías de señalización y KRAS

Este enfoque viene a complementar a los anteriores en las diferentes estrategias de terapias farmacológicas que se aplican en los tratamientos de cáncer con mutación de KRAS. En el caso del cáncer de pulmón, entre otros, las terapias dirigidas a objetivos concretos de los principales actores de las vías de señalización no han sido muy efectivas dada la gran complejidad de las redes de señalización, donde estas vías se retroalimentan por vías alternativas, minimizando así los efectos de las terapias aplicadas a los objetivos concretos. Por esta razón, muchos de los estudios actuales, buscan otras dianas terapéuticas alternativas que ayuden a entender mejor los mecanismos de señalización celular y los efectos que producen las diferentes terapias aplicadas a los pacientes, a través del análisis molecular de la expresión de los genes, que aunque no sea de forma directa, participan la activación, inhibición o retroalimentación de alguna de las vías.

En este sentido, se han identificado 8 genes que están implicados, de forma directa o indirecta en el cáncer de pulmón (NSCLC) con mutación de KRAS en las vías de señalización MAPK/ERK, que son foco de interés en muchos estudios actuales de búsqueda de dianas terapéuticas (Gil Moreno, 2018).

A continuación se detallan los genes de interés, así como su implicación den las vías de señalización celular MEK/ERK y KRAS:

- ***HES1.*** En NSCLC tratado con un inhibidor de la γ-secretasa (GSI) bloquea el crecimiento del cáncer y presentan niveles reducidos de HES1 y ERK fosforilada sin cambios en la MEK fosforilada. HES1 se une al promotor DUSP1 reprimiéndolo, que codifica una fosfatasa dual que es activa contra fosfo-ERK. Por tanto el tratamiento con GSI regula por incremento DUSP y disminuye fosfo-ERK (Maraver et al., 2012).
- ***CDCP1.*** El nivel de expresión de CDCP1 es un biomarcador importante para la señalización de RAS a través de las vías MAPK (Martinko et al., 2018)
- ***AXL.*** La pérdida de función de AXL produce un aumento de la inflamación e incluso de la autoinmunidad. Entre otras, MEK-ERK es una de las rutas de señalización activadas aguas abajo de AXL (Gay, Balaji, & Byers, 2017).
- ***CREB.*** KRAS mutante impulsa la activación de la unión al elemento de respuesta cíclico-AMP (CREB) a través de la vía de señalización RAF/MEK/ERK para forzar el crecimiento y la supervivencia de las células cancerígenas (Lee et al., 2015).
- ***LKB1.*** LKB1 (STK11) es un gen supresor de tumores que se inactiva en hasta un 30% de los NSCLC mutantes de KRAS, que promueve la diseminación agresiva y metastásica del cáncer. Provoca la resistencia a tratamientos de terapia combinada con inhibidores de MEK (Gilbert-Ross et al., 2017).
- ***Src.*** Las Src tirosina quinasas participan en la regulación de la expresión de IL-1β, IL-6 e IL-8, donde los efectos de Src están mediados por la activación de ERK (Singer at al., 2011).
- ***STAT3.*** Los tumores de pulmón con KRAS mutante presentan niveles reducidos de STAT3, que está asociado al aumento del crecimiento tumoral (Grabner et al., 2015). La respuesta heterogénea a la inhibición de MEK en estos tipos de cánceres, está influenciado entre otros por STAT3, según demuestran algunos estudios (Yoon et al., 2010).
- ***YAP1.*** Diversos estudios identifican al gen YAP1 como un mecanismo de resistencia a la terapia dirigida a RAF y MEK. El efector YAP de la vía de Hippo (codificado por YAP1) actúa como una entrada de supervivencia paralela para promover la resistencia a RAF y la terapia inhibidora de MEK (Lin et al., 2015).

En el capítulo 4 se presenta un análisis de supervivencia de pacientes con cáncer de pulmón (NSCLC) que presentaban el gen KRAS mutado y a los cuales se les habían aplicado diferentes terapias. Se ha estudiado si hay evidencia significativa estadística de influencia de la expresión de estos genes en la supervivencia de los pacientes. Se ha analizado la Supervivencia Libre de Progresión (SLP) de los pacientes, para ver si se puede determinar a través del nivel de expresión de estos genes, de qué manera pueden influir en un mejor o peor pronóstico en la progresión del cáncer tras los tratamientos y adicionalmente, en la Supervivencia Global (SGL).

Capítulo 3

3. Estudio poblacional de incidencia del cáncer de pulmón

Como ya se ha mencionado anteriormente, el cáncer en general es una de las causas principales de muerte, con 14,1 millones de casos nuevos en el 2012 y con 8.2 millones de muertes relacionadas con él, que representan más del 58 % sobre los nuevos casos de cáncer identificado al año.

Es tal el abasto del problema del cáncer, que el conocimiento estadístico sobre la incidencia de cada uno de los diferentes tipos de cáncer en la población, puede ser un buen punto de partida para orientar las líneas de investigación a realizar en los trabajos de investigación en este ámbito, focalizando aquellos *"actores principales"* que participan en los procesos de la oncogénesis de cada tipo de cáncer.

En este capítulo se va a efectuar un estudio poblacional de la incidencia, en un primer lugar, del cáncer en general a nivel mundial y del continente Europeo y posteriormente, específicamente sobre los cánceres relacionados con la vía MAPK/ERK, que son los objetivos principales del trabajo.

Tal como se ha mencionado en el apartado 1.3 donde se define las metodologías de trabajo, para realizar el estudio poblacional se ha utilizado el entorno de programación, estadístico y bioestadístico R (© Copyright 2016 RStudio Inc.; © The R Foundation). En el apartado 8.1 de anexos, se incluye la referencia del informe dinámico realizado en RMarkdown (© Copyright 2016 RStudio Inc.), que incluye el código reproducible en R de cada uno de los gráficos mostrados en este capítulo. En esta memoria se han incluido los gráficos más relevantes del estudio, pudiéndose consultar el resto en el informe completo del anexo I (apartado 8.1).

3.1 Estadística global de incidencia de Cáncer en el mundo y en Europa

Los datos analizados en este apartado se han obtenido de la Organización Mundial de la Salud (WHO - World Health Organitation, 2018) del proyecto GLOBOCAN 2012. Los indicadores incluidos hacen referencia al año 2012, dado

que a fecha de la realización del trabajo no existen datos más actualizados de la población mundial. Todos los indicadores de volúmenes hacen referencia a todos los tipos de cáncer, excluyendo el cáncer de piel no melanoma.

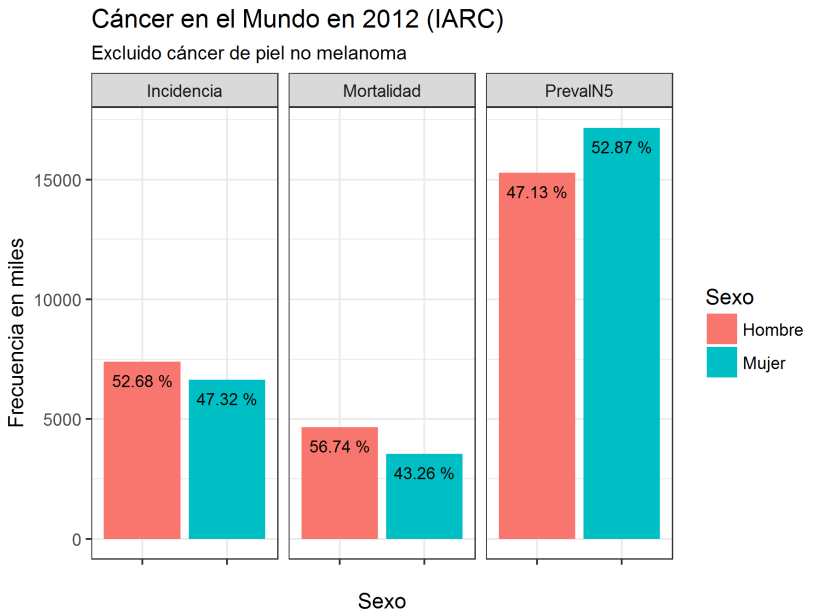
Tabla 3.1. *Distribución de Cáncer en el mundo según sexo*

			Incidencia		Mortalidad		Prevalencia
Área	Sexo	Tipos	Casos	ASR	Casos	ASR	5Años
Mundial	Hombre	All cancers excl. non-melanoma skin cancer	7410	204.9	4653	126.3	15296
Mundial	Mujer	All cancers excl. non-melanoma skin cancer	6658	165.2	3548	82.9	17159
Europa	Hombre	All cancers excl. non-melanoma skin cancer	1814	296.3	977	147.6	4508
Europa	Mujer	All cancers excl. non-melanoma skin cancer	1606	225.5	779	87.6	4571

Valores en miles (x1000).
ASR (age-standardised rate) es la tasa estandarizada por edad, una medida resumida de la tasa que tendría una población si tuviese una estructura de edad estándar, para evitar posibles diferencias entre poblaciones respecto a la edad. La tasa estandarizada por edad se calcula utilizando 10 grupos de edad (0-14; 15-39; 40-44; 45-49; 50-54; 55-59; 60-64; 65-69; 70-74; 75 +). Expresadas por 100,000 individuos por año.

En la tabla 3.1 se muestran los totales de incidencia, mortalidad y prevalencia en 5 años del cáncer a nivel mundial y de Europa.

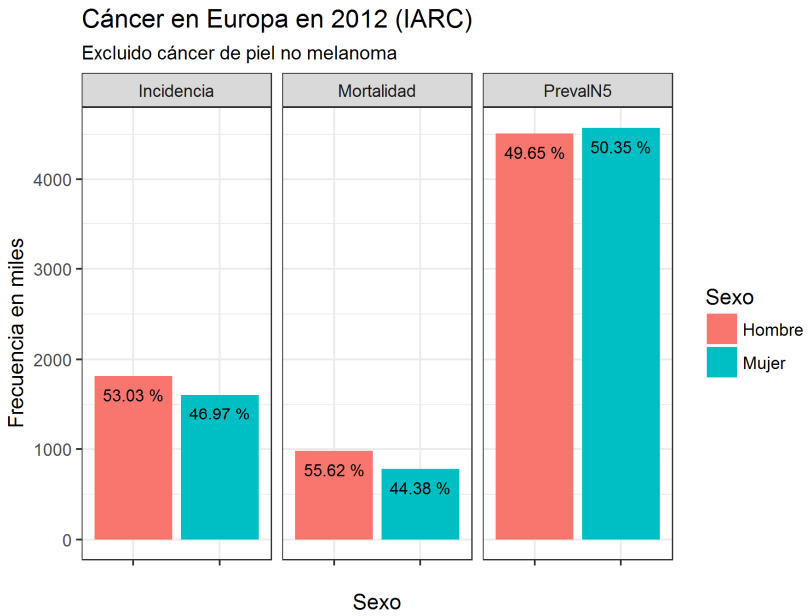
3.1.1 Incidencia del cáncer en la por Sexo



Gráfica Gr. 3.1 *Gráfica de datos de cáncer en el mundo.*

En la gráfica Gr. 3.1 se puede observar que por términos generales hay bastante paridad entre los tres indicadores, incidencia, mortalidad y prevalencia, entre hombres y mujeres. Como se puede comprobar en la tabla 3.1, los nuevos casos de cáncer son algo superiores en hombres (> 7.4 millones) respecto a las mujeres (< 6.7 millones), al igual que en la mortalidad que es más alta en hombres (> 4.6 millones) que no en las mujeres (< 3.6). La incidencia de nuevos casos de cáncer es un poco más alta en los hombres que en las mujeres, un 53% y un 47% respectivamente y respecto la mortalidad, la diferencia entre sexos aumenta con el 57% en los hombres vs al 43% en las mujeres. En cambio, la prevalencia en 5 años cambia esta tendencia, siendo un porcentaje inferior en los hombres (47%) que en las mujeres (53%).

Si se analiza la tasa estandarizada por edad (ASR) de nuevos casos de cáncer, para suponer una estructura de edades estandarizada de la población, se puede observar que las diferencias entre sexo son más significativas. A nivel mundial los hombres tienen una tasa ASR de 204.9, respecto a 165.2 en las mujeres, que representan más de un 38% de diferencia de tasa entre sexos. Respecto a la mortalidad se puede comprobar que la diferencia de este ratio entre sexos baja hasta el 34%.



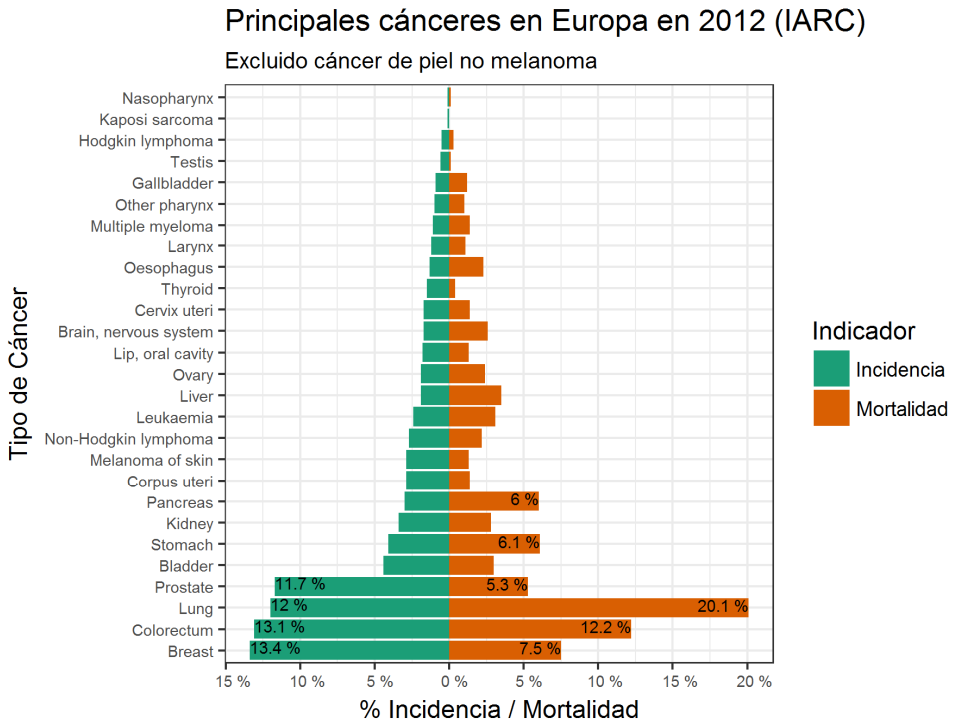
Gráfica Gr. 3.2. Gráfica de datos de cáncer en Europa.

Analizando la incidencia global del cáncer, pero esta ver a nivel de Europa (gráfica Gr. 3.2), vemos que en las distribuciones entre hombres y mujeres existe aún una mayor paridad entre los diferentes indicadores, que a nivel mundial. La prevalencia a 5 años se sitúa prácticamente al 50% en cada sexo.

Pero si se analizan las tasas ASR, se puede comprobar que éstas aumentan significativamente a las indicadas a nivel mundial. En el caso de la tasa de incidencia ASR, aumenta casi un 45%, mientras que la tasa de mortalidad aumenta aproximadamente un 17%. También se advierte que la diferencia de la tasa ASR entre hombres y mujeres en Europa aumenta en ambos indicadores, un 24% y 41% de incremento en nuevos casos y mortalidad, respectivamente y siempre superior en hombre versus las mujeres.

3.1.2 Principales cánceres en la población europea

Si se efectúa un análisis más detallado de los principales tipos de cáncer que inciden en la población, se pueden distinguir 27 cánceres con una incidencia significativa (ver gráfica Gr. 3.3), de los cuales 15 de ellos producen el 80% de la mortalidad en la población. El cáncer de pulmón encabeza el ratio de mortalidad tanto a nivel mundial, como en Europa.



Gráfica Gr. 3.3. *Gráfica de ranking de principales cánceres en Europa*

Si centramos la atención en la población europea, se puede observar que más del 50% de las muertes están provocadas por 5 cánceres: pulmón (20.1%), colón (12.2%), mama (7.5%), estómago (6.1%) y páncreas (6%), en ese orden.

3.1.3 Incidencia y mortalidad del cáncer de pulmón por regiones OMS (WHO)

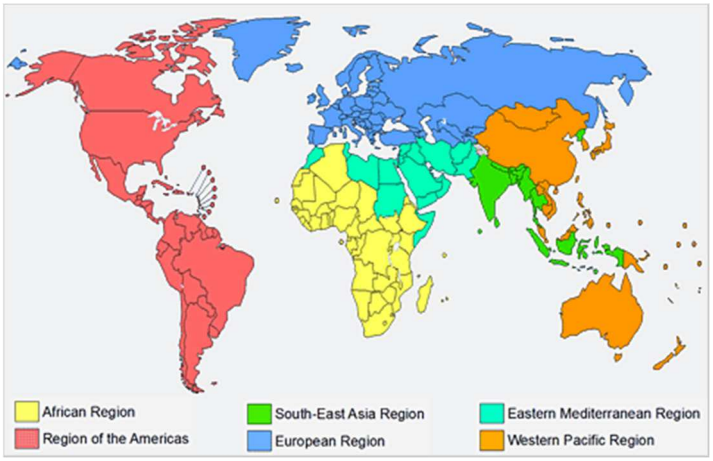
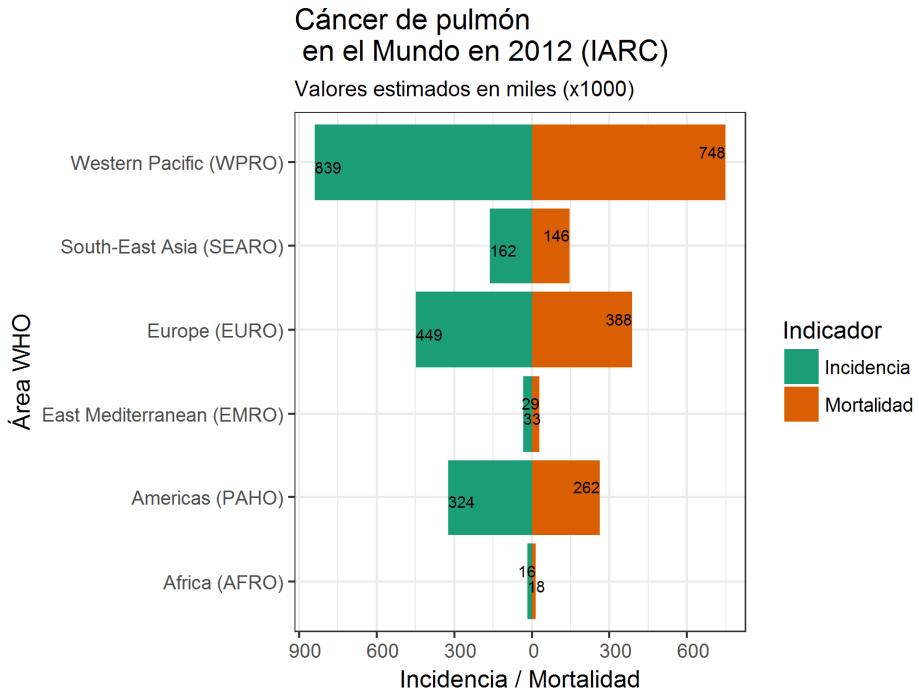


Figura F 3.1. *Regiones de la OMS (WHO)*

Tal y como se ha podido comprobar en los apartados previos, el cáncer de pulmón es el que tiene un mayor índice de mortalidad, tanto a nivel mundial como en Europa. En la gráfica Gr. 3.4 se muestra el contraste de incidencia versus mortalidad en las diferentes regiones donde la OMS recoge datos (figura F 3.1).



Gráfica Gr. 3.4. *Gráfica de Incidencia vs Mortalidad del cáncer de pulmón en mundo*

Como se puede apreciar en la gráfica Gr. 3.4, tasa de supervivencia del cáncer de pulmón es bastante baja en todas las regiones, desde un 10% en el área del sudeste de Asia al 19% de la región de Américas. Europa está en una zona intermedia con un 12% de supervivencia al cáncer de pulmón.

Dado el alto índice tanto de incidencia como de mortalidad, queda evidenciado que el cáncer de pulmón debe ser uno de los principales objetivos en los proyectos de investigación de nuevas dianas terapéuticas.

3.2 Incidencia en la población de cánceres relacionados con MAPK/ERK

Para realizar el estudio estadístico sobre la incidencia de los cánceres relacionados con los genes mutados en la vía MAPK/ERK, se han obtenido datos de muestras de tejidos con cáncer, registradas en COSMIC v84 (COSMIC v84, 2018). La base total de registros obtenida de v84 (2018) contiene 136,999 registros de muestras de tejido de diferentes tipos de cáncer con genes mutados.

3.2.1 Análisis global de frecuencia de mutaciones génicas en la vía MAPK/ERK

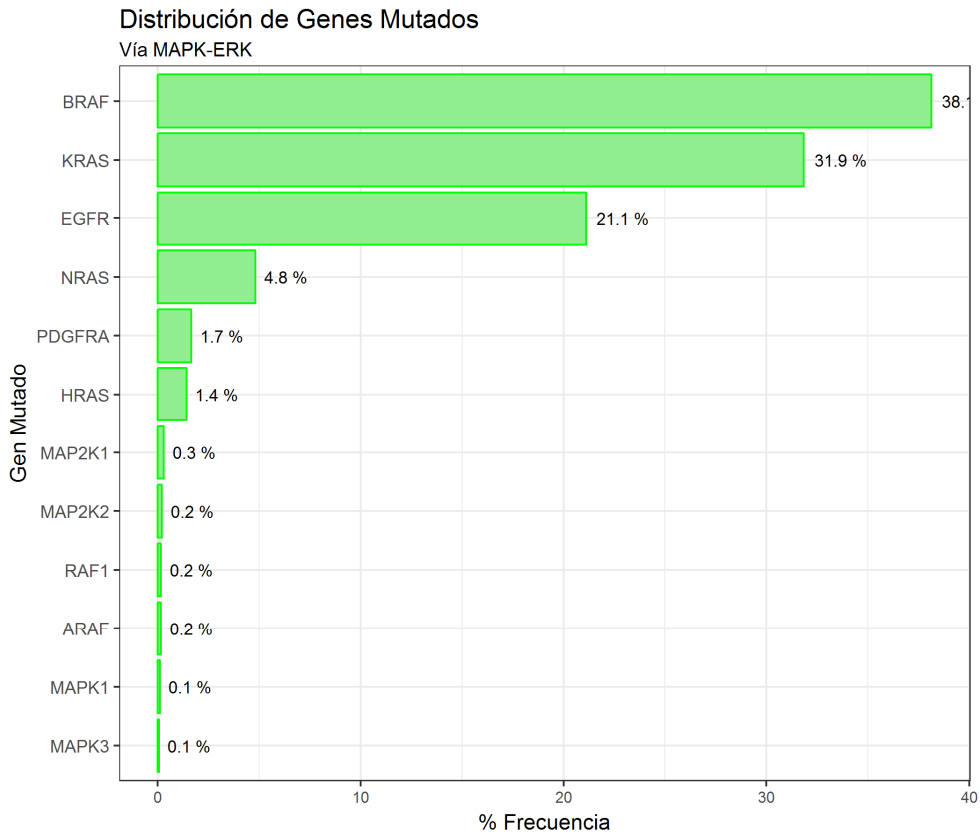
Para determinar las muestras de interés para el estudio, se han seleccionado los 12 genes de mayor interés en la vía MAPK/ERK (ARAF, BRAF, EGFR, HRAS, KRAS, MAP2K1, MAP2K2, MAPK1, MAPK3, NRAS, PDGFRA, RAF1). A partir de las mutaciones de estos genes, se han identificado un total de 38 tipos de cáncer diferentes.

En la tabla 3.2 se muestran los principales genes mutados de la vía MAPK/ERK con la frecuencia de mutación que presentan en la base de registros analizada.

Tabla 3.2 *Distribución de frecuencias de genes mutados*

Gen mutado	Frecuencia	% Frecuencia
ARAF	214	0.16
BRAF	52249	38.14
EGFR	28939	21.12
HRAS	1933	1.41
KRAS	43640	31.85
MAP2K1	407	0.30
MAP2K2	289	0.21
MAPK1	132	0.10
MAPK3	88	0.06
NRAS	6607	4.82
PDGFRA	2279	1.66
RAF1	222	0.16

Si se efectúa un ranking por frecuencia de estos genes mutados (gráfica Gr. 3.5), se puede comprobar que destacan por encima del resto, el gen BRAF, con un 38.14% de frecuencia de mutación y el gen KRAS, objetivo del trabajo, con un 31.85% de frecuencia de mutación.



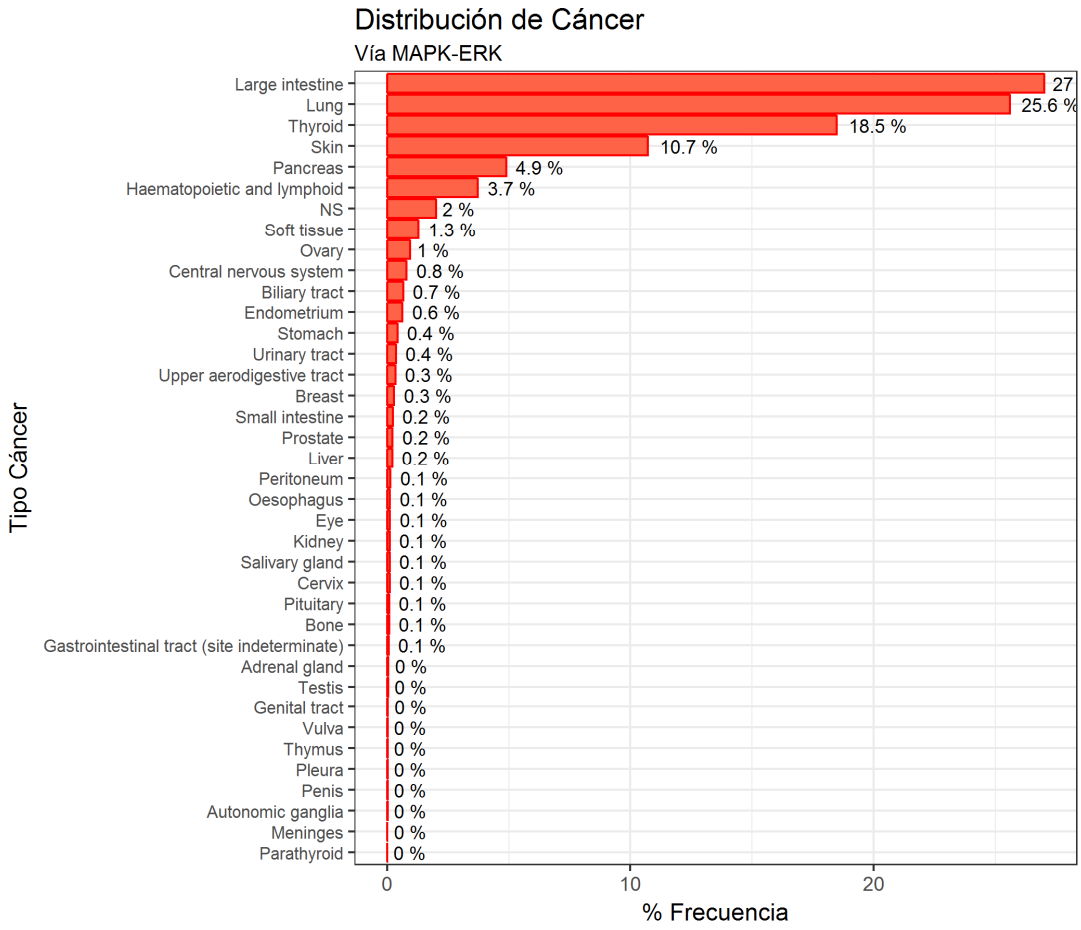
Gráfica Gr. 3.5*. Gráfica de distribución de genes mutados.*
Se puede ver el ranking de frecuencia de mutación de estos genes de mayor a menor mutación.

3.3 Comparativa de la incidencia del cáncer de pulmón con otros tipos de cáncer

Partiendo del mismo conjunto de muestras utilizado en el apartado anterior, a continuación se muestran los diferentes tipos de cáncer donde se han identificado los genes mutados en la vía MAPK/ERK. Se muestran datos y gráficas comparativas entre las frecuencias de los diferentes tipos de cáncer, conjuntamente con la frecuencia de mutación de cada gen, donde se podrá identificar el gen o genes mutación con más implicación en cada tipo de cáncer.

3.3.1 Ranking comparativo de frecuencias de tipos de cáncer con mutaciones génicas en la vía MAPK/ERK

En la gráfica Gr. 3.6 se muestra el ranking de mayor a menor frecuencia según el cáncer. Podemos observar que el cáncer de colon, el cáncer de pulmón y el de tiroides son los de mayor incidencia de mutación en esta vía, con un 27.01%, un 25.61% y un 18.5% respectivamente. Los datos detallados de frecuencia se pueden consultar en la tabla 2 el informe completo del anexo I (apartado 8.1).



Gráfica Gr. 3.6. *Distribución de frecuencias de tipos de cáncer*

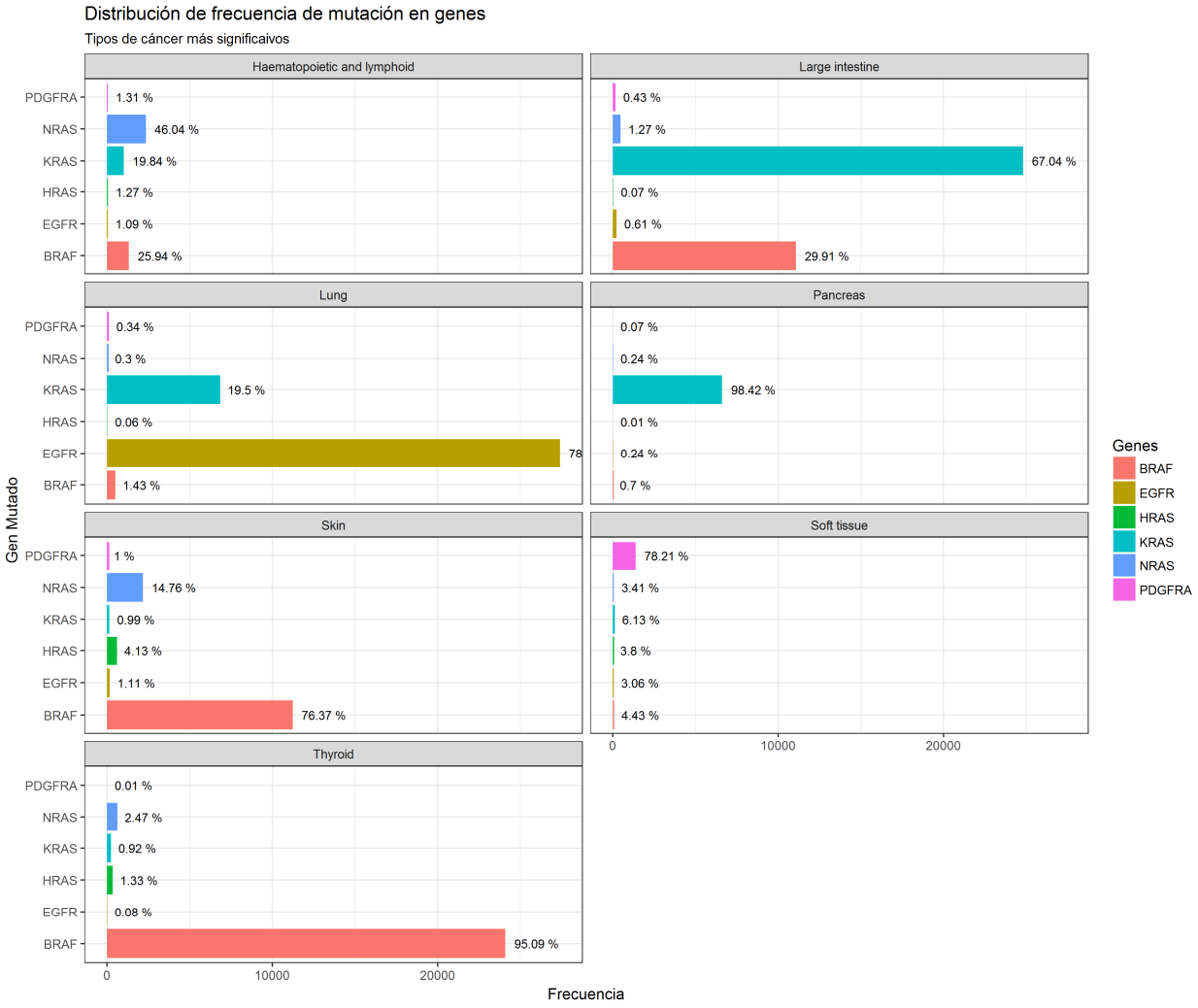
3.3.2 Análisis comparativo de implicación de mutaciones de genes según el tipo de cáncer

Para realizar los estudios comparativos con un enfoque centrado en el cáncer de pulmón y solo con el resto de cánceres más relevantes con implicación en la vía MAPK/ERK, se ha efectuado un filtraje de los genes mutados vistos en la tabla 3.2 con un porcentaje de frecuencia de aparición superior al 1%, así como de los tipos de cáncer cuya frecuencia es superior también al 1% en las muestras analizadas. Las gráficas que se muestran en adelante hacen referencia a estos tipos de cáncer. En la tabla 3.3 se pueden ver los datos numéricos del conjunto de genes mutados analizados, así como los diferentes tipos de cáncer resultado del filtraje.

Tabla 3.3. *Distribución de frecuencias de tipos de cáncer y de genes mutados*

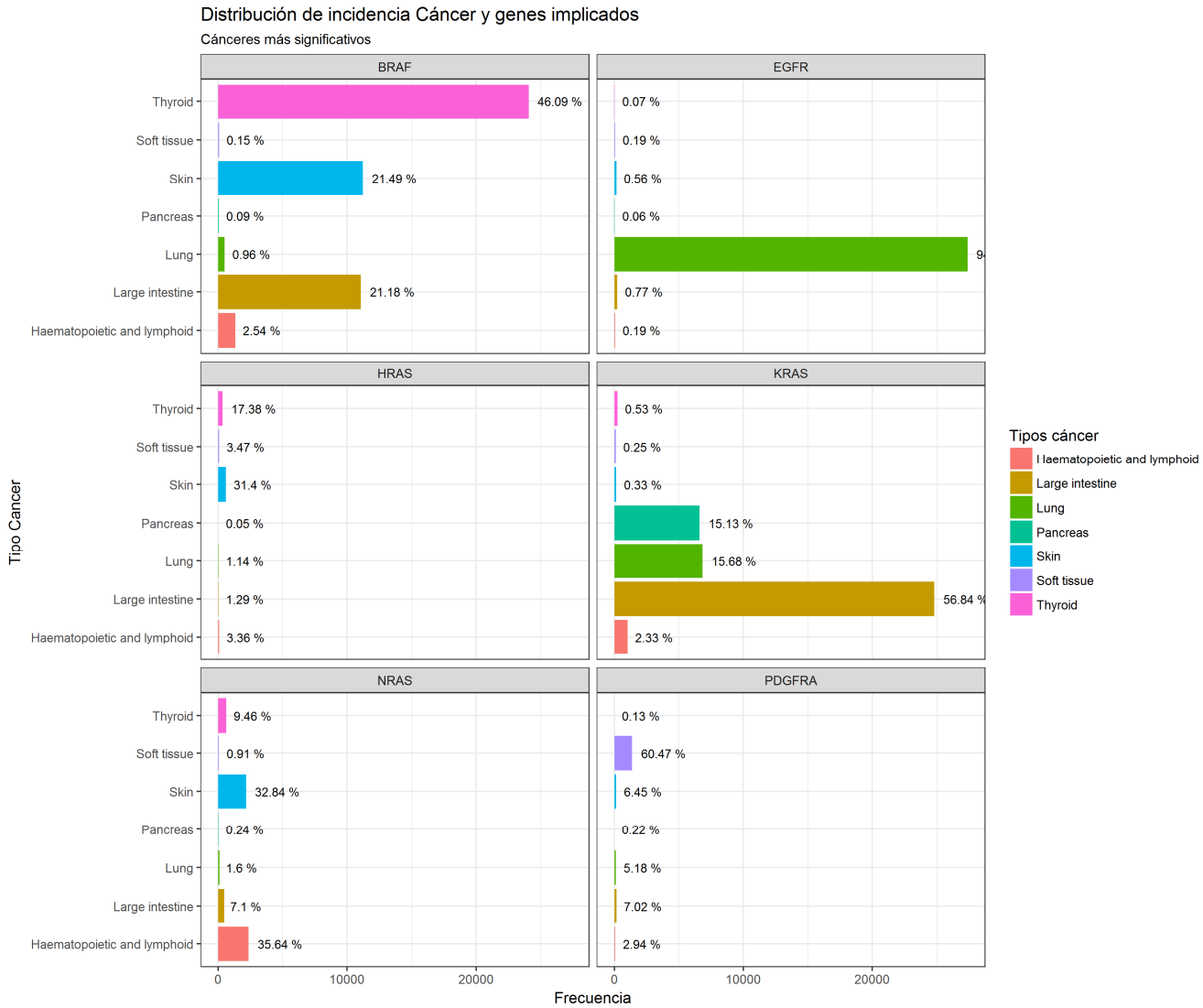
Tipo Cáncer	Gen Mutado	Frecuencia	% Gen	% Cancer
Haematopoietic and lymphoid	BRAF	1327	2.54	25.94
Large intestine	BRAF	11066	21.18	29.91
Lung	BRAF	502	0.96	1.43
Pancreas	BRAF	47	0.09	0.70
Skin	BRAF	11228	21.49	76.37
Soft tissue	BRAF	78	0.15	4.43
Thyroid	BRAF	24082	46.09	95.09
Haematopoietic and lymphoid	EGFR	56	0.19	1.09
Large intestine	EGFR	224	0.77	0.61
Lung	EGFR	27384	94.63	78.04
Pancreas	EGFR	16	0.06	0.24
Skin	EGFR	163	0.56	1.11
Soft tissue	EGFR	54	0.19	3.06
Thyroid	EGFR	20	0.07	0.08
Haematopoietic and lymphoid	HRAS	65	3.36	1.27
Large intestine	HRAS	25	1.29	0.07
Lung	HRAS	22	1.14	0.06
Pancreas	HRAS	1	0.05	0.01
Skin	HRAS	607	31.40	4.13
Soft tissue	HRAS	67	3.47	3.80
Thyroid	HRAS	336	17.38	1.33
Haematopoietic and lymphoid	KRAS	1015	2.33	19.84
Large intestine	KRAS	24804	56.84	67.04
Lung	KRAS	6843	15.68	19.50
Pancreas	KRAS	6604	15.13	98.42
Skin	KRAS	145	0.33	0.99
Soft tissue	KRAS	108	0.25	6.13
Thyroid	KRAS	232	0.53	0.92
Haematopoietic and lymphoid	NRAS	2355	35.64	46.04
Large intestine	NRAS	469	7.10	1.27
Lung	NRAS	106	1.60	0.30
Pancreas	NRAS	16	0.24	0.24
Skin	NRAS	2170	32.84	14.76
Soft tissue	NRAS	60	0.91	3.41
Thyroid	NRAS	625	9.46	2.47
Haematopoietic and lymphoid	PDGFRA	67	2.94	1.31
Large intestine	PDGFRA	160	7.02	0.43
Lung	PDGFRA	118	5.18	0.34
Pancreas	PDGFRA	5	0.22	0.07
Skin	PDGFRA	147	6.45	1.00
Soft tissue	PDGFRA	1378	60.47	78.21
Thyroid	PDGFRA	3	0.13	0.01

Para ver de forma clara los genes mutados que tienen una mayor implicación en cada tipo de cáncer, podemos ver la gráfica Gr. 3.7, donde se pueden observar los genes mutados por cada tipo de cáncer junto a la frecuencia de cada uno de ellos. En el cáncer de pulmón los dos genes más mutados son el EGFR (78.04%) y KRAS (19.5%), BRAF también aparece pero con un porcentaje de mutación muy bajo, del 1.43%.



Gráfica Gr. 3.7. *Distribución de frecuencias de mutación de genes*

Si se analiza la misma información, pero agrupando los diferentes genes mutados (gráfica Gr. 3.8), se puede observar más claramente los tipos de cáncer donde están implicados los genes de la familia RAS y concretamente KRAS.



Gráfica Gr. 3.8. *Distribución de frecuencias de mutación de genes por tipo de cáncer*

Tal como ya se mencionó en el capítulo 2, de los genes de la familia RAS, HRAS es el que menos frecuencia de mutación presenta, seguido de NRAS y KRAS como el gen que presenta una mayor frecuencia de mutación.

En la muestra analizada de tejidos KRAS ha aparecido mutado en un 56.64% de los tejidos de intestino grueso, en un 15.68% de tejidos de pulmón y en un 15.13% de tejidos de páncreas, quedando constatado que estos son los tres tipos de cáncer donde la mutación de KRAS tiene un papel relevante.

Capítulo 4

4. Estudio Bioestadístico comparativo:
Relación entre supervivencia y nivel de expresión de genes con mutación de KRAS en el cáncer de pulmón

Tal como se ha descrito en los capítulos precedentes, más del 30% del cáncer de pulmón está relacionado con las vías de señalización MAPK/ERK y con las mutaciones del gen KRAS. Además, se ha podido comprobar que uno de los retos más importantes en el ámbito de investigación del cáncer está en llegar a identificar y comprender los mecanismos de activación e inhibición de las diferentes rutas de señalización que están implicadas en el comportamiento oncogénico de las células y por tanto en la identificación de dianas terapéuticas eficaces para combatirlo. En diferentes estudios realizados, se ha evidenciado que la comprensión de estas vías se complica por la presencia de múltiples interacciones entre ellas, así como redundancias que hacen que el bloqueo o activación de las mismas no siempre produzca los efectos esperados en los tratamientos aplicados a los pacientes.

La expresión diferencial de los genes implicados en las vías de señalización puede ayudar a identificar el comportamiento y evolución del cáncer, permitiendo diseñar tratamientos más ajustados en función de sus características moleculares. En este sentido, se va a realizar un estudio bioestadístico basado en el análisis de supervivencia de una muestra de pacientes que han sido diagnosticados de cáncer de pulmón (NSCLC) con mutación del gen KRAS sometidos a diferentes terapias, para identificar, según la expresión de 8 genes implicados en las vías de señalización MAPK (identificados en el apartado 2.4.4), su influencia en la progresión del cáncer y en la supervivencia global de los pacientes. Los datos para el análisis parten de un estudio real del trabajo de la tesis doctoral de la Dra. Gil (Gil Moreno, 2018), que entre otros aspectos tratados, efectúa un análisis de supervivencia de estos genes, cuyos resultados se comparan con los obtenidos en este trabajo.

4.1 Descripción del estudio original

Abstract

El estudio original parte de una Tesis doctoral de la licenciada en medicina María de los Llanos Gil Moreno (Gil Moreno, 2018) sobre el Carcinoma de Pulmón no célula pequeña (NSCLC). A diferencia de como se ha enfocado hasta el momento el tratamiento dirigido del adenocarcinoma de pulmón con mutación en KRAS, como una enfermedad única con un mismo comportamiento, en la Tesis se plantea la hipótesis de que realmente no es así y que se trata de una enfermedad heterogénea con diferentes subtipos, que tienen comportamientos diferentes, tanto a nivel molecular, como a nivel de su evolución y por tanto, que será necesario aplicar tratamientos diferenciados a cada uno de los subtipos.

Título original de la Tesis doctoral

El título original es: "*Carcinoma de pulmón no célula pequeña, con mutación de KRAS: diferenciación y caracterización de subtipos, así como los diferentes mecanismos de resistencia, para elaboración de tratamientos dirigidos"*.

Objetivos

En la Tesis se plantean cuatro objetivos principales, que son:

1. Subdivisión en subtipos en función de las diferentes mutaciones que presenta KRAS.
2. Caracterizar por subgrupos dependiendo de la expresión génica de 8 genes analizados.
3. Diferencias entre subgrupos: característica, evolución y pronóstico.
4. Estudio de combinación de diferentes fármacos inhibidores de las vías moleculares relevantes que objetiva en el desarrollo previo de su trabajo, como posible dianas terapéuticas en 4 líneas celulares (H23, A549, H460, Calu-6).

Materiales y métodos

Se obtuvieron las muestras de pacientes diagnosticados de cáncer del pulmón célula no pequeña con mutación de KRAS, de diferentes centros previa aprobación del CEIC (Comité de ética e investigación clínica y consiguiendo los consentimientos informados de los pacientes según la ley 14/2007 de investigación biomédica. Se efectuó una valoración patológica de las muestras, tanto cuantitativa como cualitativa y de tipo de disección, para seleccionar finalmente las muestras a tratar en el estudio.

Se efectuó un análisis de las mutaciones del gen KRAS de los codones 12 y 13 mediante la técnica HRM (High Resolution Melting) de Applied Biosystems.

El análisis de la expresión génica de ARNm de los genes de interés, se realizó a partir de la obtención de RNA de las muestras primarias, mediante una lisis celular con una solución de EDTA 0.1M, SDS 2%, tris 0.5M y proteinasa K (20 mg/ml). El cDNA se sintetiza usando un Kit basado en la encima de retrotranscripción MMLV (200u/µL)/RNASE OUT (20u/µL), realizando controles

de técnica: controles de extracción y calibradores. La amplificación se realizó mediante PCR a Tiempo Real introduciendo por triplicado 2.5 µLde ADNc obtenido de las muestras, con un sistema de detección ABI Quanti studio Flex de Applied Biosystems, utilizando sondas y cebadores específicos para cada uno de los genes a analizar. La cuantificación se realiza a partir de los valores de fluorescencia que se transforman mediante un software al valor Ct, que es el número de ciclos necesarios para que haya un aumento de fluorescencia significativo respecto a la señal de base. Se realiza una validación de sondas y cebadores, mediante un control de calidad in silico.

El valor relativo de expresión se calculó utilizando el método de Ct comparativo usando β-actin como endógeno y RNA comercial como calibrador (Liver and Lung; Stratagene, La jolla, CA USA). Los estadísticos calculados para la expresión génica fueron: promedio de las tres réplicas, dCt de cada muestra para cada uno de los genes analizados, que se define como el promedio de la diferencia entre el gen problema con el gen endógeno, como δCt o dCt, la media de dCt de los calibradores y por último el ddCt para las muestras y los calibradores, calculado según la fórmula: $ddCt = 2 - dCt_{gen} - dCt$ (liver-lung). Adicionalmente, se realizan experimentos in Vitro para valorar el comportamiento de las líneas celulares tumorales H23, A549, H460 y Calu-6, frente a inhibidores antitumorales.

A nivel estadístico se realiza un estudio de SLP, supervivencia libre de progresión, que se definió como el tiempo desde el inicio del tratamiento de primera línea hasta la progresión o muerte y un estudio de la SG, supervivencia global, definida como el tiempo desde el inicio del tratamiento de la primera línea hasta la muerte. Se utilizaron modelos de regresión de riesgos proporcionales de Cox para los análisis multivariantes. Todo el análisis estadístico se realizó con el software SPSS v15 (Statistical Package for Social Sciencies).

4.2 Analisis bioinformático y bioestadístico

4.2.1 Objetivos del proceso de análisis

El tercer objetivo global planteado en el trabajo se basa en el estudio de las implicaciones de KRAS en el cáncer de pulmón. Para su consecución, se propone efectuar un análisis bioinformático y bioestadístico, que evidencie a través de la expresión de 8 genes implicados en las líneas tumorales con mutación de KRAS, su influencia en la supervivencia de los pacientes libres de progresión tumoral (SLP) y adicionalmente en la supervivencia global (SGL), realizando posteriormente una comparativa de los resultados obtenidos respecto al estudio original.

Los datos de las muestras del estudio han sido cedidos por la Dra. María de los Llanos Gil, con fines únicamente de realización del presente TFM, para compartir y contrastar posteriormente los resultados obtenidos. Todo el proceso se efectuará en el entorno de programación y plataforma estadística R (© The R Foundation), generando un código reproducible de todas las funciones utilizadas. Como resultado final se ha generado un informe dinámico en RMarkdown con todos los procesos de análisis realizados, junto con los resultados gráficos y numéricos obtenidos. El informe completo se anexa al TFM como uno de los

productos finales del trabajo, para su consulta. En este capítulo se reflejan los procesos y datos más relevantes obtenidos en el estudio que forman parte del informe de análisis.

El código R implementado para todas las funciones realizas en el informe están incluidas en el apartado de "*Anexo de código R…"* del informe global de análisis (anexo II, apartado 8.2).

4.2.2 *Pipeline* para el proceso de análisis

Se propone el siguiente Pipeline de trabajo para la realización del análisis:

- Obtención, preparación y exploración de los datos:
  - Preparación de los datos
  - Exploración de los datos
  - Análisis de normalidad y correlación
- Proceso de análisis de Supervivencia
- Conclusiones.
- Comparativa de resultados con estudio original

4.2.3 Proceso de análisis

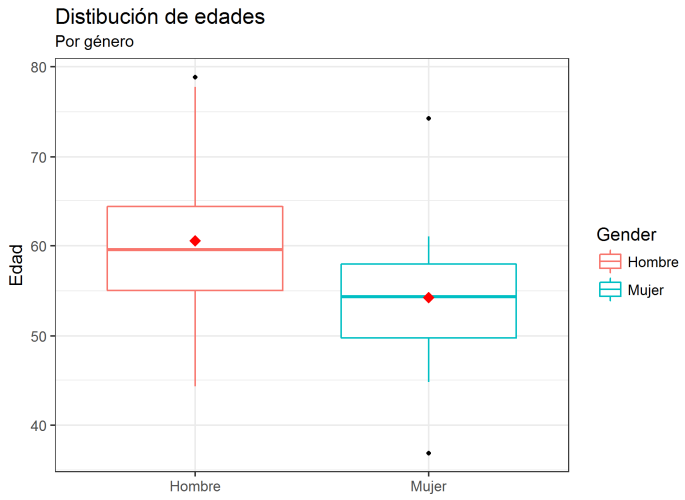
El informe completo del análisis bioestadístico realizado se adjunta como anexo a esta memoria del TFM, con el título*: Análisis Supervivencia de Cáncer NSCLC – Pacientes con mutación KRAS* (anexo II, apartado 8.2).

4.2.3.1 Obtención, preparación y exploración de los datos

El fichero original (*BaseFinalAnalisisEstadísticoKRAS_DraGil.csv*) con los datos del estudio, contiene 32 registros de muestras de adenocarcinomas de pulmón, cada uno con 79 variables que corresponden a los datos clínicos recogidos para el análisis, a los datos de seguimiento de los tratamientos administrados a los pacientes y a los datos sobre los cálculos de expresión génica realizados a los genes objetivo del estudio original. Se han descartado 2 registros por falta de los datos básicos necesarios para la realización del análisis de supervivencia, por lo que finalmente la muestra contiene 30 pacientes.

4.2.3.1.1 Preparación y exploración de los datos

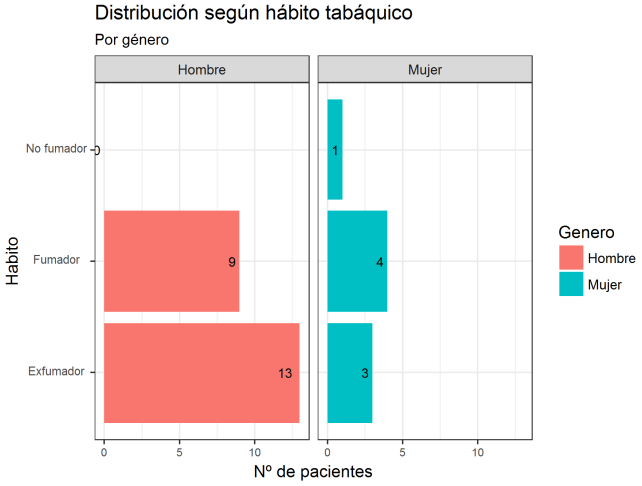
Para la realización del análisis, de los datos completos del estudio se han seleccionado únicamente las variables requeridas para la obtención del análisis de supervivencia. El conjunto de datos final de trabajo contiene las 30 muestras seleccionadas, de las cuales se han dejado únicamente las 48 variables que son más relevantes para el proceso. Estas variables se han transformado según su tipología para realizar posteriormente una mejor exploración de los datos. El detalle completo de los datos se puede ver en la tabla 1 del informe completo.

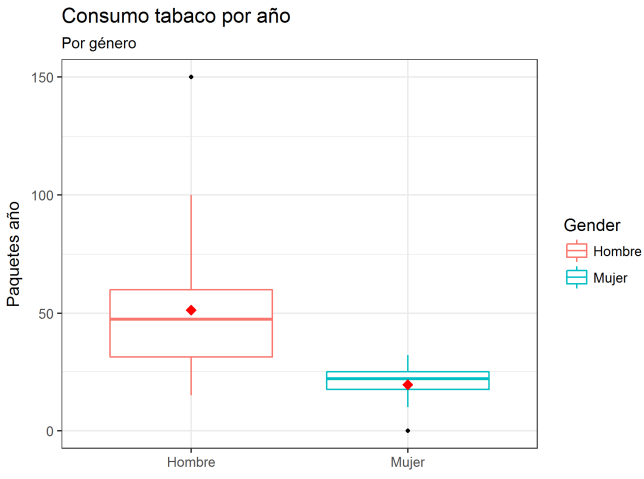


Gráfica Gr. 4.1. *Distribución de las edades de los pacientes por género.*

El conjunto de datos contiene una mayoría de hombres con un 73.33%, respecto al 26.67% de mujeres. La edad media global de los pacientes es de 58.89 años.

La edad media de las mujeres es inferior a la de los hombres, de 54 y 61 años respectivamente. La mediana global de edad de la muestra es de 58.61 años. En la gráfica Gr.4.2 se presenta la distribución de pacientes de la muestra respecto su hábito de consumo de tabaco, clasificados en función del sexo. Se puede observar, que prácticamente todos los pacientes eran fumadores o exfumadores, solo hay un caso que nunca había fumado. Aunque la muestra ya determina su mayor porcentaje de hombres que mujeres, se puede ver que la proporción de fumadores hombres es el doble que el de mujeres, aunque el número de exfumadores proporcionalmente es mucho mayor en hombres que en las mujeres. En referencia al consumo de tabaco, si se cuenta el consumo de los fumadores y el consumo que tenían los exfumadores, la media del consumo está en 44 paquetes por año. Si se analiza por género, en la gráfica Gr.4.3 se puede ver que en base al nivel de consumo de paquetes de tabaco por año, el hombre consume mucho más que la mujer, con un consumo medio de unos 51 paquetes por año en hombres y unos 22 paquetes por año en mujeres.

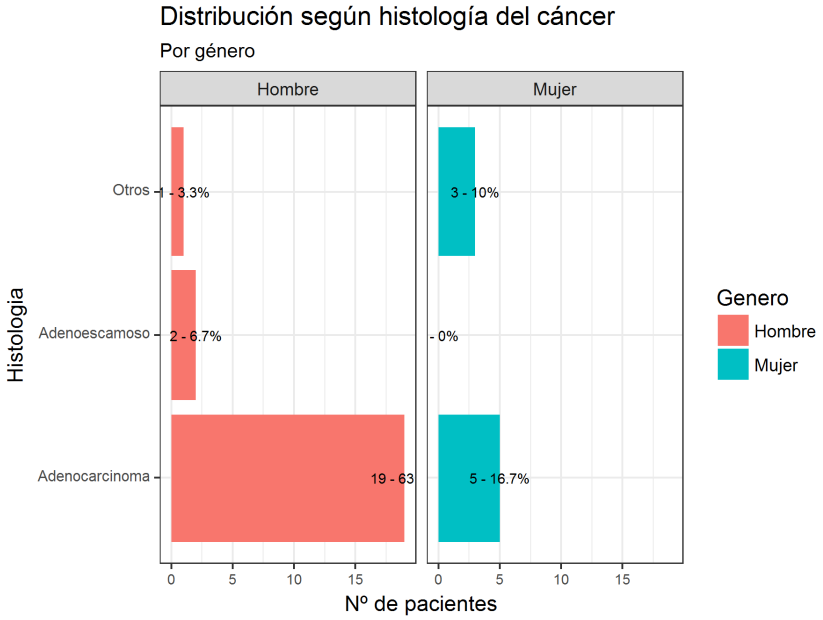




Gráfica Gr. 4.2 *Hábito tabáquico*

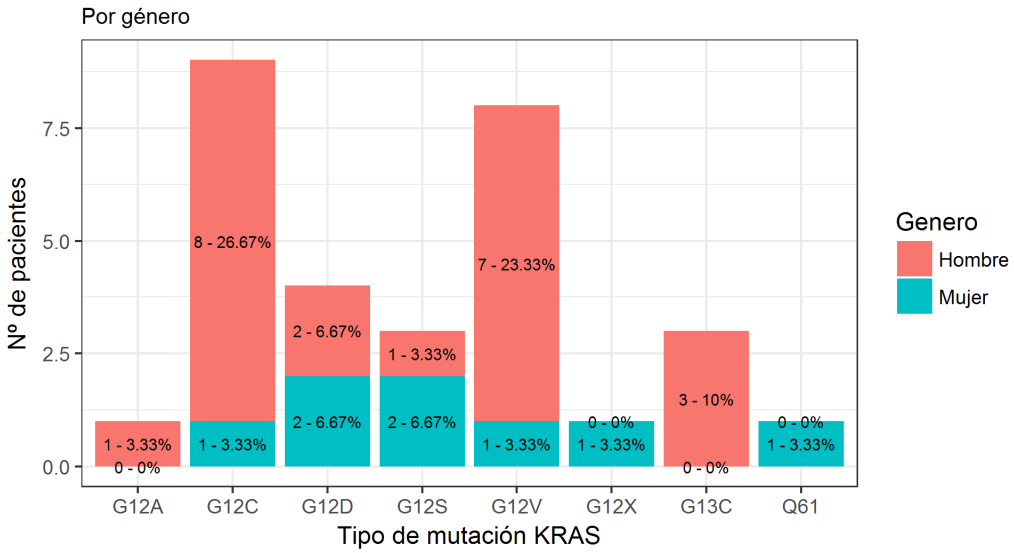
Gráfica Gr. 4.3 *Consumo de tabaco*

Respecto a la histología de los tejidos analizados de la muestra (gráfica Gr.4.4), el 80% eran Adenocarcinomas y el 6.7% son Adenoescamosos. En ambos sexos el Adenocarcinoma es el más frecuente, aunque en proporción ha sido más habitual en los hombres que en las mujeres.



Gráfica Gr. 4.4 Distribución según histología del cáncer**.**

Respecto a las características moleculares de las muestras, el análisis de la mutación de KRAS que se ha realizado, muestra que los tres tipos de mutación más frecuentes son G12C, G12V y G12D con el 70%, seguido de G12S y G13C con el 10% cada una de ellas (ver gráfica Gr.4.5).



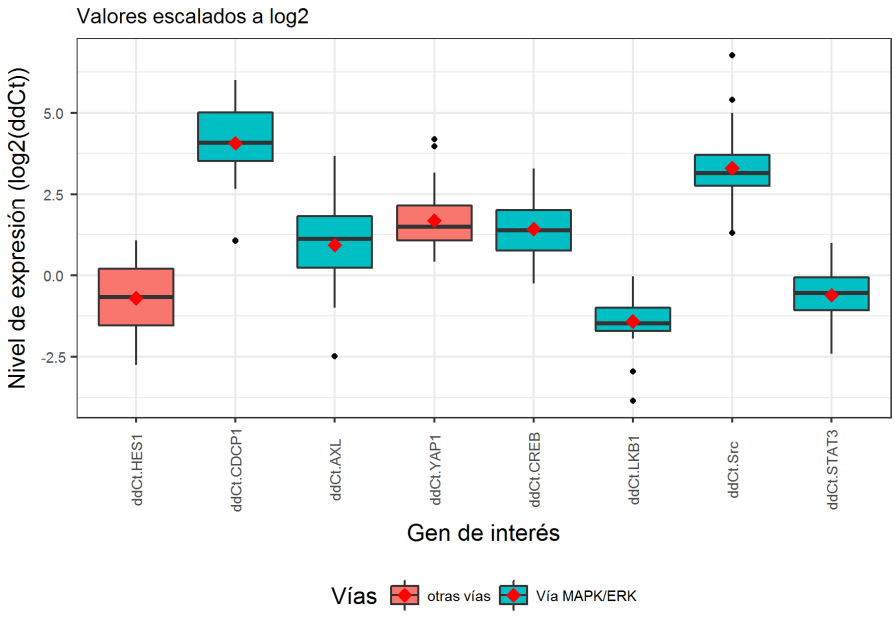
Gráfica Gr. 4.5 *Distribución según mutación KRAS*

Tal como se apuntó en el apartado 2.4.4, se identificaron 8 genes que están implicados en los mecanismos de activación de las vías de señalización celular donde KRAS aparece mutado y que pueden generar que las terapias aplicadas a los pacientes no tengan los resultados esperados. Estos 8 genes parecen estar

implicados como activadores o inhibidores de las vías MAP quinasas con KRAS mutado. Hay 6 de ellos que están relacionados más directamente con las vías de señalización MAPK/ERK y otros 2 que aunque no directamente y dada la redundancia de las vías acaban influyendo también en la funcionalidad de esta vía.

Al explorar las variables de las expresiones de estos genes, se puede comprobar que los datos recogidos para cada muestra y gen no son completos en todos los casos. Presentan una gran variabilidad de valores entre la expresión de los diferentes genes y además en cada uno de ellos existen valores omitidos (NA's), posiblemente debido a la calidad de las muestras analizadas en cada caso. Los genes con mayor número de valores omitidos son el HES1 y el LKB1 y los que menos valores omitidos presentan han sido el gen CDCP1 y Scr. En el informe global puede verse un resumen estadístico de las variables de la expresión de cada gen. La expresión esta medida en base al cálculo ddCt de expresión.

En el siguiente gráfico de cajas, Gr.4.6, se muestra la distribución de los valores de expresión de los 8 genes, a los cuales se les ha realizado un re-escalado mediante una transformación logarítmica para obtener una escala que permita una fácil comparación para ver si guardan cierta homogeneidad. Como se puede comprobar los datos de expresión no son homogéneos entre los diferentes genes y por tanto la media parece no ser una buena medida de resumen entre las diferentes expresiones. Se han agrupado los genes en base a su implicación directa o no en la vía de señalización de MAPK/ERK. En cualquier caso, se confirma que hay divergencia en la distribución de los valores de expresión en los diferentes genes.



Gráfica Gr. 4.6. *Nivel de expresión génica*

4.2.3.1.2 Análisis de normalidad y correlación

Adicionalmente se ha verificado la normalidad o no de los valores de expresión, para la aplicación de funciones estadísticas paramétricas o no paramétricas. Tanto los histogramas, como las gráficas qqplot realizadas para cada variable de expresión (gráficas Gr. 8 y Gr. 9 del informe de análisis completo en anexo II, apartado 8.2) indican la no normalidad de los valores de expresión.

Para validar estadísticamente este aspecto, se ha realizado un test de normalidad de Shapiro-Wilks de cada variable, donde se confirma que todas las distribuciones de la expresión de los genes tienen un $p - valor < 0.05$ por lo que se descarta la distribución normal, a excepción del gen STAT3 que tiene un $p - valor > 0.05$ por lo que en este caso no se puede descartar la normalidad (tabla 4.1).

Tabla 4.1. *Test de Shapiro-Wilk de expresiones génicas*

gen	W	p.value
Src	0.51155	0.00000
YAP1	0.61403	0.00000
AXL	0.68917	0.00001
CREB	0.82433	0.00096
LKB1	0.86663	0.01026
CDCP1	0.92757	0.05355
STAT3	0.92661	0.06439
HES1	0.91484	0.09086

Otro aspecto relevante en la exploración de los datos de expresión de los genes, ha sido determinar si existe o no correlación entre ellas. Para ello y dado que se ha verificado la no normalidad de los datos de expresión, se ha optado por efectuar un test no paramétrico de correlación de Spearman, implementado en la función *cor.test()* de R. Dado que la implementación de esta función es para realizarlas únicamente entre parejas de datos, se ha definido la función *rs.test()* en R para obtener los resultados de correlación entre todas las parejas posibles de los 8 genes estudiados. Los valores numéricos de correlación, así como los valores de significancia entre los genes se detallan en la tabla 4 del informe de análisis completo (anexo II, apartado 8.2). Hay evidencia significativa estadísticamente al 0.05 o menos de que hay correlación entre los genes LKB1 y CREB con mayor correlación, seguida de los genes YAP1 y SRC, de YAP1 y STAT3, y con menor fuerza de correlación entre STAT3 y HES1 (ver gráfica Gr.4.7).

Gráfica Gr. 4.7.*Test de correlación de Spearman.*
El grado de correlación vienen identificado por el área e intensidad del color de los círculos mostrados, cuanto mayor es su área y mayor intensidad de color, mayor es el valor de correlación rho, el color azul indica correlación positiva y el rojo correlación negativa. El grado de significación viene marcado por los asteriscos (,**,***).*

4.2.3.2 Proceso de análisis de Supervivencia

Respecto al proceso de análisis de supervivencia, es necesario indicar que debido a las características especiales del tipo de estudio realizado sobre pacientes de cáncer de pulmón con mutación en KRAS, a los que se les han aplicado diferentes terapias de tratamiento en momentos diferentes en temporalidad, no existe un periodo de análisis concreto definido para realizar el estudio de supervivencia, como sería la forma más convencional de realizar este tipo de estudios.

Los datos recogidos para el estudio fueron obtenidos de diferentes fuentes, por lo que cada paciente pertenece a un periodo de tratamiento determinado diferente, que varía desde el año 2008 al año 2016, momento en el cual fueron diagnosticados y tratados. Por tanto, el estudio de supervivencia se efectúa relativo al inicio del tratamiento de primera línea una vez diagnosticada la primera metástasis de cada paciente o la fecha del diagnóstico de la primera metástasis o en su defecto el diagnóstico del cáncer, en función del dato disponible.

De todos los pacientes que inicialmente formaban parte del estudio, se descartaron aquellos sobre los cuales el análisis molecular del tejido tumoral no obtuvo los mínimos exigidos de calidad para obtener resultados correctos. Tal como se indicó en el apartado de descripción de los datos, el conjunto final de

muestras está compuesto por 30 pacientes, de los cuales se descartan adicionalmente 2 pacientes por no disponer datos clínicos suficientes para poder establecer el periodo de tratamiento para calcular el estadístico de supervivencia. Se verá a continuación que en los diferentes test y modelos de regresión se descartan siempre estos dos registros de pacientes, quedando la muestra final de análisis con 28 pacientes.

En la tabla 4.2 se muestra el conjunto de datos final para el análisis de supervivencia. La variable *ObjSLP* contiene los valores *time* par el análisis, donde el signo + indica que ese paciente está *censurado*.

4.2.3.2.1 Supervivencia Libre de Progresión (SLP)

Para calcular la variable *time* para el análisis de *Supervivencia Libre de Progresión (SLP)*, se ha tenido en cuenta como fecha de inicio del estudio, la fecha de inicio del tratamiento que se administró a los pacientes en primera línea. El tiempo se ha calculado en meses. De todos los pacientes con datos clínicos disponibles, se identificaron 3 pacientes de los cuales no se disponía de la fecha de inicio del tratamiento, ni la fecha de diagnóstico de la primera metástasis, por lo que se tomó la fecha del diagnóstico de cáncer como fecha de inicio (según se adoptó en el estudio original). Como fecha de producción del evento, se tomó la fecha de éxitus del paciente o la fecha en que se determina que después de finalizar la línea del tratamiento correspondiente, se constata que la progresión del cáncer continúa. Si se revisan los datos iniciales de la muestra, al inicio de cada línea de tratamiento (*X2nd.line... X5th.line*) se codifica la progresión del cáncer. Todos los códigos diferentes de *A22* indican una progresión en la enfermedad. En caso de progresión se toma la fecha de inicio del tratamiento de la siguiente línea, como fecha de ocurrencia del evento. Hay tres casos en que se desconoce la fecha de progresión o éxitus, por lo que se toma la última fecha de seguimiento disponible, como fin de periodo y marcando estos pacientes como censurados (en la tabla 4.2, la variable *ObjSLP* de los pacientes censurados tiene un símbolo +).

Para el proceso de análisis de supervivencia se han seguido tres estrategias diferentes para el análisis de Supervivencia Libre de Progresión (SLP) y para el análisis de Supervivencia Global (SGL), que son las siguientes:

- Análisis de la supervivencia en global sin estratificación mediante la estimación de *Kaplan-Meier.*
- Análisis de la supervivencia en global estratificada por género y por tipo de mutación de KRAS mediante la estimación de *Kaplan-Meier*.
- Ajuste de *modelos de regresión de Cox* para SLP y SLG: Univariable y Multvariable.

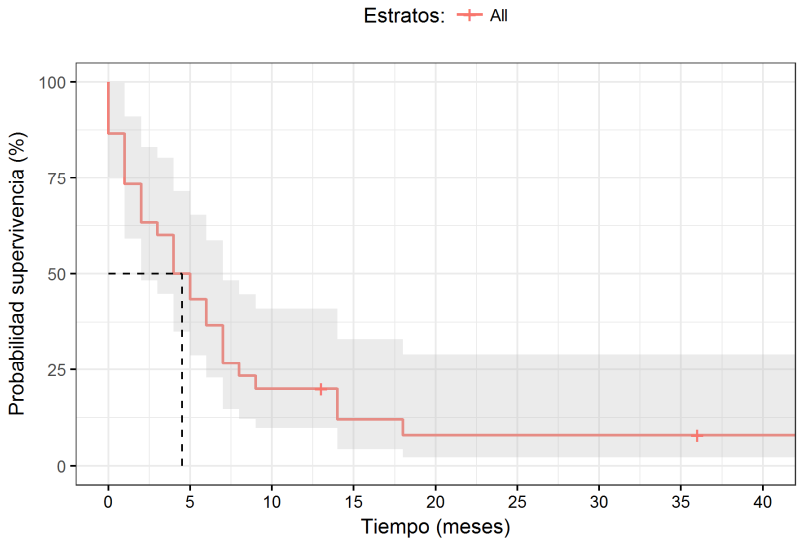
A continuación se detallan algunos de los resultados en cada uno de las estrategias implementadas en el análisis. En el informe de "*Analisis de supervivencia…"* completo (anexo II, apartado 8.2) se pueden consultar los detalles de cada análisis realizado.

Tabla 4.2. *Datos de la muestra para el Análisis de Supervivencia*

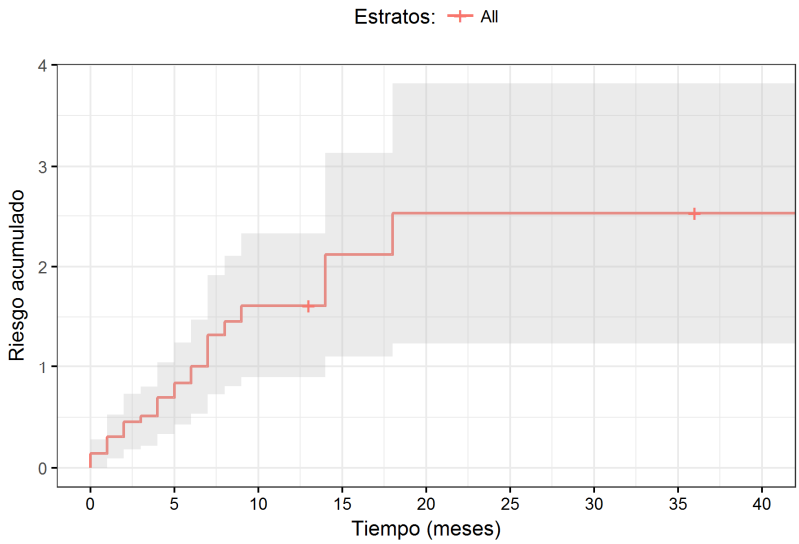
Nº.KRAS	Gender	Smoking.historyFac	PA.YEAR	Mutacion	ddCt.HES1	ddCt.CDCP1	ddCt.AXL	ddCt.YAP1	ddCt.CREB	ddCt.LKB1	ddCt.Src	ddCt.STAT3	ObjSLP	ObjSGL
kras 3	Hombre	Exfumador	60	G12V	0.35	16.81	1.06	1.37	1.34	0.36	3.68	0.46	36+	36+
kras 7	Hombre	Fumador	15	G12V	0.29	43.75	0.66	3.47	0.84	0.07	10.71	0.55	1	1
kras 10	Hombre	Exfumador	60	G12C	0.51	6.37	1.36	1.95	1.94	0.31	6.87	0.43	4	4
kras 11	Hombre	Exfumador	150	G12C	NA	28.67	3.49	1.69	1.31	0.26	2.62	0.19	1	2+
kras 13	Hombre	Exfumador	35	G12C	0.63	15.79	3.69	3.15	2.50	0.35	8.58	1.98	14	18
kras 14	Hombre	Exfumador	60	G12D	1.01	32.03	1.17	4.68	2.59	0.47	9.46	0.84	13+	13+
kras 15	Mujer	No fumador	0	G12D	1.22	36.03	2.87	3.04	3.64	0.61	11.72	0.97	3	23+
kras 16	Mujer	Exfumador	24	G12V	2.09	8.12	2.07	2.14	4.18	0.98	20.41	0.67	7	7
kras 17	Hombre	Exfumador	30	G12C	NA	32.47	1.10	18.32	7.64	NA	19.72	1.23	7	26+
kras 19	Hombre	Exfumador	36	G12V	1.28	7.22	1.15	4.39	3.00	0.32	7.22	1.14	4	30
kras 20	Hombre	Fumador	100	Otras	0.34	12.83	1.25	3.04	3.47	0.43	12.92	0.26	5	4+
kras 22	Hombre	Fumador	35	Otras	NA	21.13	5.59	NA	NA	NA	42.26	NA	14	16
kras 23	Hombre	Exfumador	20	G12C	0.45	15.90	1.70	4.81	4.45	0.64	17.40	0.68	2	37+
kras 24	Mujer	Fumador	20	G12C	0.77	2.06	2.15	2.78	2.05	0.38	7.68	0.63	6	19
kras 28	Mujer	Fumador	10	G12D	0.20	8.23	2.55	8.97	2.32	0.30	5.07	1.30	4	8
kras 31	Hombre	Exfumador	30	G12C	1.73	28.08	NA	4.48	3.04	NA	7.90	0.73	8	26
kras 32	Hombre	Exfumador	50	G12V	0.15	16.01	3.79	1.77	3.79	0.36	2.45	0.37	0	8
kras 33	Hombre	Exfumador	80	G12C	NA	64.06	NA	15.73	NA	NA	107.73	0.91	6	9
kras 34	Hombre	Fumador	45	G12D	NA	44.06	12.84	NA	NA	NA	NA	NA	0	0+
kras 36	Mujer	Fumador	25	Otras	NA	17.16	2.18	1.33	4.54	0.96	3.76	0.29	44+	44+
kras37	Mujer	Exfumador	32	Otras	NA	19.44	NA	NA	NA	NA	10.06	1.00	1	2
kras39	Mujer	Fumador	20	Otras	NA	42.55	3.71	2.39	1.60	NA	6.03	0.53	2	3
kras 42	Hombre	Fumador	60	G12V	0.70	36.28	3.90	2.35	1.79	0.36	9.20	0.74	9	15
kras43	Hombre	Exfumador	45	G12V	NA	NA	NA	NA	NA	NA	31.81	NA	2	2
kras 44	Mujer	Exfumador	25	Otras	NA	NA	NA	NA	NA	NA	NA	NA	1	1
kras 47	Hombre	Fumador	60	G12V	0.63	8.29	2.43	2.08	1.47	0.37	8.35	0.69	5	5
kras 48	Hombre	Fumador	15	G12C	1.07	14.74	2.79	2.35	9.81	0.85	6.59	0.92	7	18+
kras 49	Hombre	Fumador	25	Otras	0.21	2.10	1.45	4.68	NA	0.13	13.84	0.58	0	0
kras 50	Hombre	Fumador	50	Otras	NA	24.27	0.50	2.08	1.34	NA	7.42	0.36	18	38
kras 52	Hombre	Exfumador	70	Otras	1.48	14.84	0.18	2.02	4.36	0.29	12.05	1.07	0	0

4.2.3.2.1.1 Análisis de la SLP global sin estratificación

En la tabla 6 del informe de análisis completo (anexo II, apartado 8.2) se muestra el resultado detallado de los valores obtenidos del análisis SLP. En las gráficas Gr.4.8 y Gr.4.9 mostradas a continuación, se muestran las curvas de supervivencia y de riesgos acumulados respectivamente. Como se puede observar, la mediana de SLP (50% probabilidad de supervivencia) es de *4.5* meses. La media calculada de SLP resultante es de *8.2* meses.



Gráfica Gr. 4.8 *Curva de Supervivencia SLP*



Gráfica Gr. 4.9. *Riesgo acumulado SLP*

4.2.3.2.1.2 Análisis de la SLP global estratificado por género y tipo de mutación KRAS

Para verificar si el sexo de los pacientes o el tipo de mutación del gen KRAS tienen alguna influencia sobre el tiempo de SLP, adicionalmente se ha efectuado

un análisis de supervivencia estratificado por sexo y por tipo de mutación en KRAS, esta última según las mutaciones más frecuentes y el resto (G12C, G12D, G12V y otras). Los detalles del resultado numérico y gráficos de ambos análisis se pueden consultar en el informe completo. Aunque en ambas curvas de supervivencia, por grupo de sexo y mutación, no parecen mostrar diferencias significativas de la SLP para los grupos comparados, con la función *survdiff()* del paquete *survival* (Therneau, 2015), se ha efectuado un test de diferencias entre los resultados de las estimaciones de supervivencia de ambos grupos, para determinar si éstas eran o no significativas estadísticamente. En ambos casos el *p-valor* obtenido ha sido < *0.05*, por lo que se descarta la hipótesis de diferenciación entre los estratos analizados.

4.2.3.2.1.3 Ajuste de modelos de regresión de Cox Univariable para SLP

Uno de los objetivos finales del trabajo era determinar si la expresión diferencial de los genes analizados molecularmente en el estudio, tenían influencia sobre la supervivencia de los pacientes. Esto significa que era necesario efectuar un modelo de regresión para ver si la expresión de los diferentes genes podía ser un riesgo o un factor influyente para la supervivencia de los pacientes. Para ello, se ha ajustado un *modelo de regresión de Cox*, que nos proporciona la significancia del riesgo proporcional para cada una de las covariables de expresión de los genes en la supervivencia de los pacientes. En el apartado "*Ajuste de un modelo de regresión de Cox para SLP" del i*nforme de análisis completo (anexo II, apartado 8.2) se describen los fundamentos el modelo de Cox, así como la validación de que el modelo sea una representación adecuada de los datos, verificando el supuesto de riesgos proporcionales.

En primer lugar, se efectúa un modelo de regresión univariable, es decir, un modelo de regresión con cada una de las covariables (variables de expresión génica) de forma independiente y sin estratificación o agrupación. Posteriormente, en base al resultado obtenido, se realiza un modelo de regresión multivariable buscando el modelo más ajustado y utilizando una estrategia de descarte progresivo de variables hasta encontrar el mejor modelo, en base a la significación global del mismo, el índice de concordancia *C* y la significación de cada covariable. Por último y también en función de los resultados obtenidos, se efectúa un modelo de regresión estratificado en base a diferentes criterios, con la finalidad de buscar un modelo más ajustado que apoye o descarte la hipótesis de influencia de la expresión de algún gen sobre la supervivencia de los pacientes del estudio.

Para efectuar el modelo de ajuste de regresión de Cox de cada covariable con el tiempo de supervivencia (variable *ObjSLP*), se define la función *fit.cox()* en R que implementa la ejecución de la función *coxph()* del paquete *survival* (Therneau, 2015) de R, para las variables de conjunto de datos a analizar. En la misma función se han calculado los puntos de corte óptimos (valor en *CutP* de la tabla 4.3) obtenidos mediante la función *cutp()* del paquete *survMisc* (Dardis, 2016) de R.

Tabla 4.3. *Resultados del modelo de regresión de Cox Univariable.*

	n	n.eventos	beta	HR (95% CI for HR)	wald.test	p.value	CutP
ddCt.HES1	19	17	-0.2439	0.7836 (0.2902-2.1155)	0.23	0.6303	0.35
ddCt.CDCP1	28	25	0.0016	1.0016 (0.9711-1.033)	0.01	0.9192	16.81
ddCt.AXL	25	22	0.1494	1.1611 (0.9444-1.4274)	2.01	0.1563	1.25
ddCt.YAP1	25	22	0.0191	1.0193 (0.9315-1.1153)	0.17	0.6779	1.69
ddCt.CREB	23	20	0.0105	1.0106 (0.8259-1.2365)	0.01	0.9186	4.54
ddCt.LKB1	20	17	-2.2456	0.1059 (0.0076-1.4838)	2.78	0.0955	0.35
ddCt.Src	28	25	0.0024	1.0024 (0.9845-1.0207)	0.07	0.7922	10.06
ddCt.STAT3	26	23	0.0656	1.0678 (0.432-2.6394)	0.02	0.8870	0.53

En la tabla 4.3 se puede comprobar que de forma individual no parece que la expresión de ninguno los genes influya de forma significativa en la supervivencia libre de progresión. El $p - valor > 0.05$ en todas las variables de expresión. En base a estos resultados, se optó por investigar si a diferentes niveles de expresión de cada gen podía haber evidencia significativa de que a esos niveles la expresión fuera un factor influyente en el tiempo de supervivencia de los pacientes. Para ello, se efectuaron diferentes niveles de estratificación de las variables de expresión, ajustando nuevamente un modelo de regresión de Cox de cada covariable estratificada en cada criterio. Inicialmente se probaron tres estratificaciones de los valores de expresión siguiendo criterios de proporcionalidad por percentil igual para todos los genes y posteriormente se aplicó una estratificación basada en el cálculo del mejor punto de corte para cada covariable en el modelo de Cox, obtenido según el método propuesto por Contal y O'Quigley (Contal & O'Quigley, 1999) implementado mediante la función *cutp()*. Finalmente, los criterios de estratificación implementados han sido los siguientes:

- Estratificación por los percentiles 33 y 66, efectuando una agrupación factorizada de la expresión en los niveles de expresión *Baja* ($\leq percentil\ 33$), *Media* ($> percentil\ 33\ y \leq percentil\ 66$) y *Baja* ($> percentil\ 66$).
- Estratificación por niveles de expresión *Baja-Media* ($\leq percentil\ 66$) y *Alta* ($> percentil\ 66$).
- Estratificación por dos cuantiles de expresión, cuantil inferior a 25 etiquetados como *<Q25* y superiores o iguales a 25, etiquetados como *>=Q25*.
- Mejores puntos de corte de la expresión de cada gen, calculado con la función *cutp()* del paquete *survMisc* de R (Therneau, 2015).

Para los cálculos de los percentiles indicados, se ha utilizado la función *quantile()* de R.

Estratificación según expresión Alta, Media y Baja

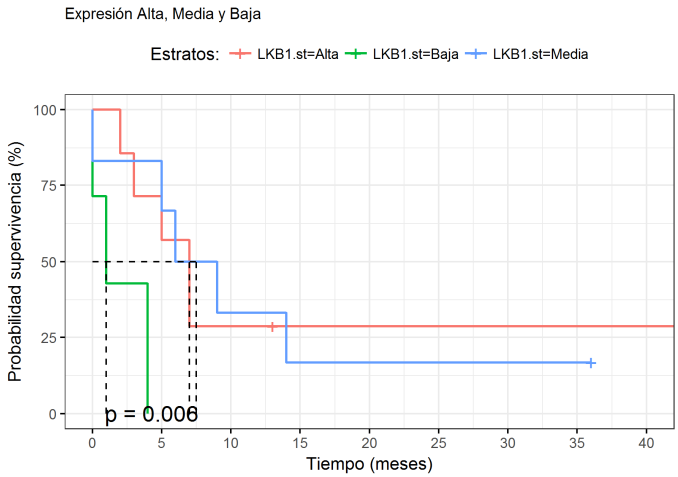
Una vez realizado el modelo de ajuste de regresión de Cox con la expresión estratificada de los diferentes genes, se puede comprobar (datos en rojo en la tabla 4.4) que el gen *LKB1* muestra evidencia de significación al *5%* en el estrato de expresión *Baja* respecto a la expresión *Alta*, con un $p - valor = 0.0071$ y con un *test de Wald* también significativo con un $p - valor = 0.0139$. Según este modelo de estratificación para el gen *LKB1* se clasificaron 7 pacientes como expresión *Alta*, 6 pacientes como expresión *Media* y 7 pacientes como expresión

Baja, 10 pacientes presentaban valores omitidos (NA's) para la expresión de este gen.

Tabla 4.4. *Modelo de Cox Univariable Estratificado (Alta, Media y Baja)*

Variable	n	n.eventos	wald.test	p.value	strat	beta	HR (95% CI for HR)	Pr(>|z|)
HES1.st	19	17	0.36	0.834	HES1.stBaja	0.3374	1.4013 (0.4118-4.7683)	0.5892
					HES1.stMedia	0.0199	1.0201 (0.3234-3.2172)	0.973
CDCP1.st	28	25	3.66	0.1602	CDCP1.stBaja	0.2756	1.3173 (0.5035-3.4464)	0.5743
					CDCP1.stMedia	-0.9321	0.3937 (0.1196-1.2963)	0.1252
AXL.st			1.27	0.53	AXL.stBaja	-0.6	0.5488 (0.1912-1.5749)	0.2646
					AXL.stMedia	-0.1735	0.8408 (0.3067-2.3044)	0.736
YAP1.st	25	22	0.96	0.6192	YAP1.stBaja	-0.5789	0.5605 (0.1756-1.789)	0.3282
					YAP1.stMedia	-0.1952	0.8227 (0.3068-2.2062)	0.6982
CREB.st	23	20	0.35	0.838	CREB.stBaja	-0.2198	0.8027 (0.2778-2.3197)	0.6848
					CREB.stMedia	-0.3235	0.7236 (0.2401-2.1808)	0.5655
LKB1.st	20	17	8.56	0.0139	LKB1.stBaja	2.0587	7.8361 (1.7509-35.0695)	0.0071
					LKB1.stMedia	-0.032	0.9685 (0.277-3.3864)	0.96
Src.st	28	25	2.12	0.347	Src.stBaja	-0.5302	0.5885 (0.2166-1.5986)	0.2984
					Src.stMedia	-0.6651	0.5142 (0.1988-1.3299)	0.1701
STAT3.st	26	23	1.13	0.5693	STAT3.stBaja	-0.5313	0.5878 (0.2051-1.6844)	0.3226
					STAT3.stMedia	-0.3853	0.6802 (0.2505-1.8474)	0.4497

La curva de supervivencia para el gen *LKB1* según esta estratificación se muestra en la gráfica Gr.4.10, donde se puede apreciar el cambio significativo en la supervivencia de los pacientes con un nivel de expresión *Baja* de este gen versus al nivel *Alto*.



Gráfica Gr. 4.10. *Curva de Supervivencia LKB1 estratificada*

Estratificación según expresión Alta, Baja-Media

Tal como reflejan los resultados de los diferentes modelos de Cox univariables con esta estratificación (tabla 4.5), no existe evidencia de significación al 5% de ninguno de los genes analizados.

Tabla 4.5. *Modelo de Cox Univariable Estratificado (Alta, Baja-Media)*

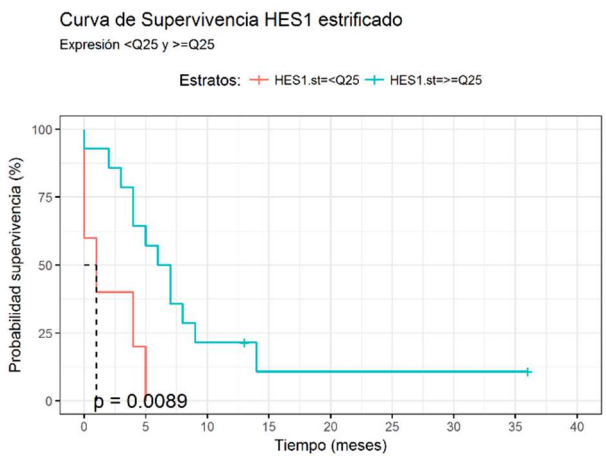
Variable	n	n.eventos	wald.test	p.value	strat	beta	HR (95% CI for HR)
HES1.st	19	17	0.08	0.7785	HES1.stBaja-Media	0.1467	1.158 (0.4167-3.2181)
CDCP1.st	28	25	0.39	0.5335	CDCP1.stBaja-Media	-0.2706	0.7629 (0.3255-1.7882)
AXL.st	25	22	0.78	0.3772	AXL.stBaja-Media	-0.392	0.6757 (0.2831-1.6129)
YAP1.st	25	22	0.58	0.447	YAP1.stBaja-Media	-0.3489	0.7055 (0.287-1.7341)
CREB.st	23	20	0.32	0.5698	CREB.stBaja-Media	-0.2691	0.7641 (0.3021-1.9326)
LKB1.st	20	17	1.37	0.2417	LKB1.stBaja-Media	0.6275	1.8728 (0.6551-5.3542)
Src.st	28	25	2.06	0.1507	Src.stBaja-Media	-0.6059	0.5456 (0.2387-1.2467)
STAT3.st	26	23	1.07	0.3005	STAT3.stBaja-Media	-0.4543	0.6349 (0.2687-1.5004)

Estratificación según expresión según expresión <Q25 y >=Q25

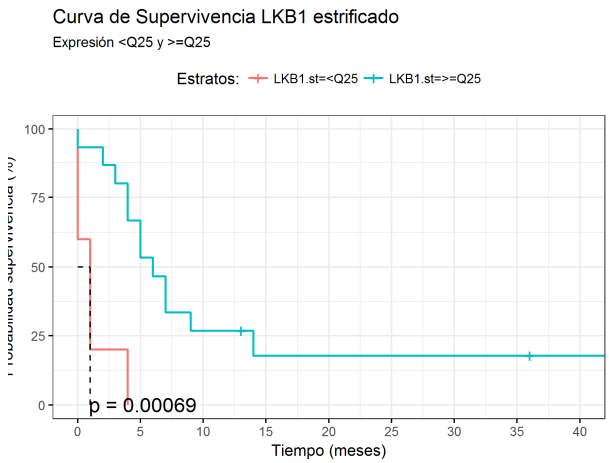
Tabla 4.6. *Modelo de Cox Univariable Estratificado <Q25 y >=Q25*

Variable	n	n.eventos	wald.test	p.value	strat	beta	HR (95% CI for HR)
HES1.st	19	17	6.15	0.0131	HES1.st>=Q25	-1.5425	0.2138 (0.0632-0.7234)
CDCP1.st	28	25	1.95	0.1622	CDCP1.st>=Q25	-0.6765	0.5084 (0.1969-1.3126)
AXL.st	25	22	0.4	0.5258	AXL.st>=Q25	0.3318	1.3935 (0.5-3.884)
YAP1.st	25	22	0.33	0.5659	YAP1.st>=Q25	0.3316	1.3932 (0.4492-4.321)
CREB.st	23	20	0.03	0.8622	CREB.st>=Q25	0.0929	1.0974 (0.3844-3.1324)
LKB1.st	20	17	9.78	0.0018	LKB1.st>=Q25	-2.1108	0.1211 (0.0323-0.4549)
Src.st	28	25	0.66	0.4164	Src.st>=Q25	0.4176	1.5182 (0.5546-4.1565)
STAT3.st	26	23	1.92	0.1659	STAT3.st>=Q25	0.7909	2.2053 (0.7204-6.7509)

Como resultado de los modelos de Cox analizados en esta estratificación (tabla 4.6), se puede ver que hay dos covariables que tienen evidencia de significación al 5% (en rojo), el gen *HES1* y el gen *LKB1*, en su estrato de >=*Q25* respecto a <*Q25*, con $p-valores\ de\ 0.0131\ y\ 0.0018$, y un $HR\ de\ 0.21, 0.12$, respectivamente. La distribución de los pacientes según esta estratificación para los dos genes significativos fue la siguiente: en el gen *HES1* se han clasificado 5 pacientes como <*Q25* y 14 pacientes como >=*Q25*, en el caso del gen *LKB1* se han clasificado 5 pacientes <*Q25* y 15 pacientes como >=*Q25*, el resto hasta los 30 registros de cada gen han sido valores omitidos (NA's). En las gráficas Gr.4.11 y Gr.4.12 se muestran las curvas de supervivencia de la expresión estratificada de los genes *HES1* y *LKB1*.



Gráfica Gr.4.11



Gráfica Gr.4.12

Estratificación según expresión con mejores puntos de corte

En la tabla 4.7 se muestra el resultado de la estratificación según los mejores puntos de corte de cada gen, donde se pueden ver los valores de cada uno y la distribución de los pacientes.

***Tabla 4.7**. Estratificación: Mejores puntos de corte*

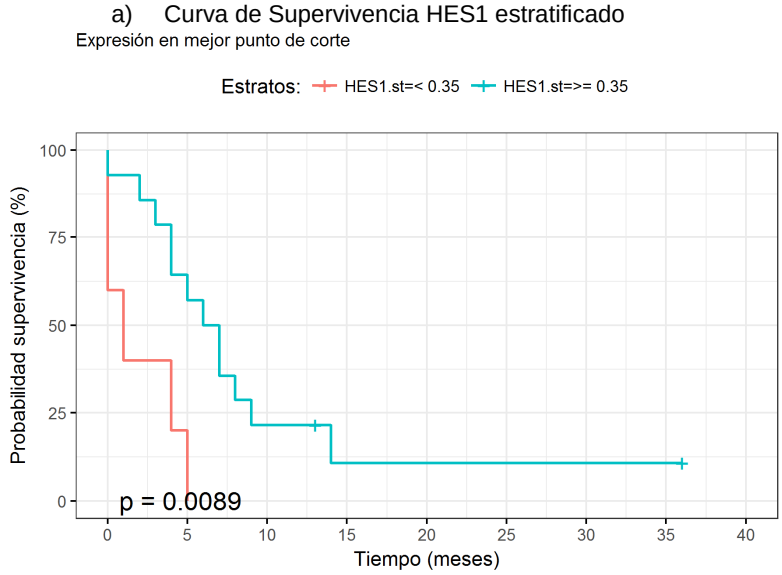
HES1.st	CDCP1.st	AXL.st	YAP1.st	CREB.st	LKB1.st	Src.st	STAT3.st
< 0.35 : 5	< 16.81 :13	< 1.25 : 7	< 1.69 : 2	< 4.54 :20	< 0.35 : 7	< 10.06 :16	< 0.53 : 7
>= 0.35 :14	>= 16.81 :15	>= 1.25 :18	>= 1.69 :23	>= 4.54 : 3	>= 0.35 :13	>= 10.06 :12	>= 0.53 :19
NA's :11	NA's : 2	NA's : 5	NA's : 5	NA's : 7	NA's :10	NA's : 2	NA's : 4

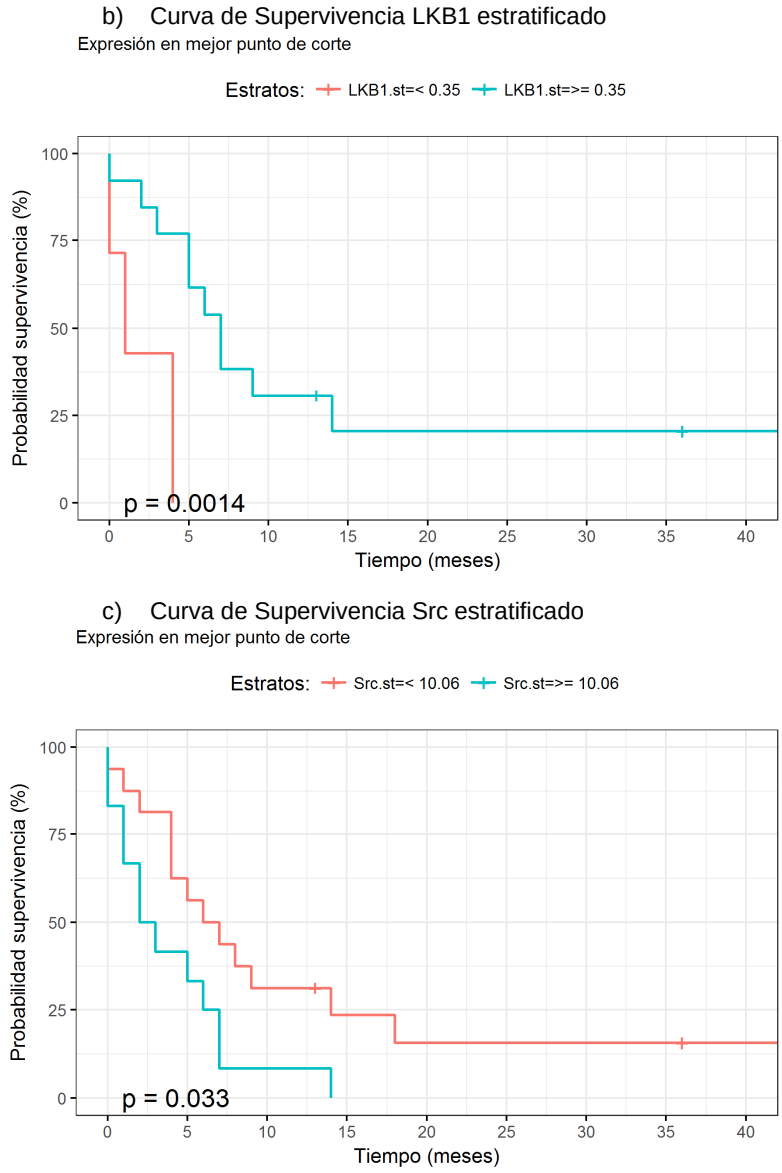
En rojo, en la tabla 4.8, se indican los modelos de Cox que han resultado significativos al 5% con esta agrupación. Podemos ver que éstos son los genes: *HES1*, *LKB1* y *Src*, con un $p - valor\ de$ 0.0131, 0.0034, 0.0346, y un $HR\ de$ 0.21, 0.13, 2.43, respectivamente.

***Tabla 4.8.** Modelo de Cox Univariable estratificado con mejores puntos de corte*

Variable	n	n.eventos	wald.test	p.value	strat	beta	HR (95% CI for HR)
HES1.st	19	17	6.15	0.0131	HES1.st>= 0.35	-1.5425	0.2138 (0.0632-0.7234)
CDCP1.st	28	25	3.65	0.056	CDCP1.st>= 16.81	-0.8201	0.4404 (0.1899-1.0211)
AXL.st	25	22	1.45	0.2281	AXL.st>= 1.25	0.6246	1.8675 (0.6762-5.1574)
YAP1.st	25	22	0	0.9978	YAP1.st>= 1.69	19.681	352637377.3775 (0-Inf)
CREB.st	23	20	1.28	0.258	CREB.st>= 4.54	-0.8479	0.4283 (0.0985-1.8616)
LKB1.st	20	17	8.56	0.0034	LKB1.st>= 0.35	-2.0732	0.1258 (0.0314-0.5046)
Src.st	28	25	4.46	0.0346	Src.st>= 10.06	0.8862	2.4258 (1.0663-5.5186)
STAT3.st	26	23	1.92	0.1659	STAT3.st>= 0.53	0.7909	2.2053 (0.7204-6.7509)

Vemos las curvas de supervivencia de los tres genes identificados como significativos, cada uno según la estratificación de su mejor punto de corte en las gráficas Gr.4.13 a, b y c.





Gráfica Gr. 4.13. *Curvas de Supervivencia*

Viendo las curvas y valores obtenidos en el modelo ajustado según los mejores puntos de corte, se puede concluir que los niveles inferiores a *0.35* de *ddCt* en la expresión del gen *HES1* nos indican un mal pronóstico, dado que la mediana de supervivencia se sitúa en 1 mes, en comparación a los 6.5 meses de mediana en valores de expresión superiores o iguales a este rango de *0.35 de ddCt*. De forma equivalente, en el caso del gen *LKB1* se puede comprobar que valores inferiores también a *0.35 de ddCt*, tienen un peor pronóstico versus los niveles superiores o iguales a *0.35 de ddCt*, que pasan de una supervivencia de 1 mes, en los niveles de expresión bajos, a 7 meses en los niveles de expresión más altos. En cambio, si se observa la curva de supervivencia del gen *Src*, el comportamiento es el contrario, los niveles de expresión inferiores a *10.06* de *ddCt* tienen un mejor pronóstico que los valores de expresión superiores a este rango. La supervivencia mediana de los niveles bajos es de 6.5 meses, en contraste con los 2.5 meses de mediana en los niveles altos *(>=10.06*).

Todos los valores de expresión que han resultado significativos se han validado con un test de Log-Rank de comparación por pares, cuyo resultado confirma la significación de todos ellos.

4.2.3.2.1.4 Ajuste de modelos de regresión de Cox Multivariable para SLP

Para ver si se puede predecir la función de supervivencia en base a la expresión de genes analizados, se ha planteado la realización de un modelo de regresión de Cox multivariable. En este caso, se deberá considerar que se excluirán las muestras en las cuales la expresión de alguno de los genes no esté informada (valor NA), por lo que el modelo solo contemplará la *n* mínima de las muestras que tienen valor de expresión para todos los genes, por lo que posiblemente su resultado no sea concluyente. Por otra parte, verificar el principio de linealidad de las variables de expresión con una *n* tan baja de muestras es difícil, ya que un solo valor outlier puede alterar el resultado de un test de linealidad, por lo que se asumirá la linealidad de las covariables, como ejercicio de prueba del modelo. Como principio mínimo para la aplicación del modelo de Cox Mutivariable, se ha validado la suposición de riesgos proporcionales entre las covariables de expresión, validación básica para la aplicación de un modelo de Cox.

En un primer modelo de ajuste se introdujeron todas las covariables, obteniendo que solo tres de ellas fueron significativas a *0.05, CREB*, *LKB1* y *Src*, con un $p - valor = 0.0049, 0.0027\ y\ 0.020$, respectivamente, pero el test de razón de verosimilitud (*Likekihook*) indicó un $p - valor > 0.05$, por lo que el modelo no podía ser considerado como un modelo representativo de los datos. Con este resultado, se buscó un modelo con un mejor ajuste, efectuando diferentes combinaciones de covariables de entrada y verificando los estadísticos de significación, tanto de las covariables, como del modelo global. Como estrategia para la selección de variables, se optó por desarrollar una función en R, *fit.cox2()*, que generaba los modelos de Cox partiendo del modelo completo con todas la covariables y eliminando una a una las covariables hasta dejar las tres significativas del modelo completo (*CREB*, *LKB1* y *Src*). Se registraron internamente los estadísticos conseguidos en cado modelo, para determinar así el modelo más ajustado. En las tablas 4.9 y 4.10 se muestran los resultados obtenidos.

En las tablas se puede comprobar que el modelo de regresión que utiliza las covariables de la expresión de los genes *CREB*, *LKB1*, *Src* y *STAT3* es significativo en dos de los tres test de ajuste (línea en color rojo), en el de razón de verosimilitud (Likekihook) y en el test de Wald. De estas cuatro variables explicativas, el modelo indica que las variables *CREB*, *LKB1* y *Src* son significativas, con un $p - valor < 0.05$, y la variable *STAT3* tiene un $p - valor > 0.05$, por lo que no es significativa al 5%. El *índice de Concordancia* del modelo $C = 0.7562$, que es un valor bastante aceptable. Como se puede ver, solo se han podido utilizar 19 muestras de las 30 de conjunto de datos inicial.

Tabla 4.9. *Variables significativas de cada modelo de regresión de Cox*

Modelo	Var.Significativas	p.valores
HES1+CDCP1+AXL+YAP1+CREB+LKB1+Src+STAT3	CREB, LKB1, Src	0.0049, 0.0027, 0.0199
CDCP1+AXL+YAP1+CREB+LKB1+Src+STAT3	CREB, LKB1, Src	0.0152, 0.0079, 0.0295
AXL+YAP1+CREB+LKB1+Src+STAT3	CREB, LKB1, Src	0.0155, 0.0081, 0.0474
YAP1+CREB+LKB1+Src+STAT3	CREB, LKB1	0.0161, 0.0074
CREB+LKB1+Src+STAT3	CREB, LKB1, Src	0.0146, 0.0059, 0.0464
CREB+LKB1+Src	CREB, LKB1	0.0294, 0.0114
CDCP1+YAP1+CREB+LKB1+Src+STAT3	CREB, LKB1, Src	0.016, 0.0067, 0.0422
CDCP1+CREB+LKB1+Src+STAT3	CREB, LKB1, Src	0.0144, 0.0054, 0.0354
CDCP1+CREB+LKB1+Src	CREB, LKB1, Src	0.03, 0.0103, 0.0466
AXL+CREB+LKB1+Src+STAT3	CREB, LKB1, Src	0.0138, 0.0063, 0.0406
AXL+YAP1+CREB+LKB1+Src	CREB, LKB1	0.0285, 0.0118

Tabla 4.10. *Significación global de cada modelo Multivariable*

Modelo	n	n.eventos	Conc.	likelih.pval	wald.pval	score.paval
HES1+CDCP1+AXL+YAP1+CREB+LKB1+Src+STAT3	17	15	0.7656	0.0879	0.2217	0.1091
CDCP1+AXL+YAP1+CREB+LKB1+Src+STAT3	19	16	0.7938	0.0569	0.192	0.1591
AXL+YAP1+CREB+LKB1+Src+STAT3	19	16	0.775	0.0522	0.1417	0.1263
YAP1+CREB+LKB1+Src+STAT3	19	16	0.7688	0.04	0.0893	0.0943
CREB+LKB1+Src+STAT3	19	16	0.7562	0.0216	0.0486	0.0546
CREB+LKB1+Src	19	16	0.6688	0.0214	0.0705	0.0862
CDCP1+YAP1+CREB+LKB1+Src+STAT3	19	16	0.775	0.0572	0.1388	0.1403
CDCP1+CREB+LKB1+Src+STAT3	19	16	0.7625	0.0342	0.0835	0.0903
CDCP1+CREB+LKB1+Src	19	16	0.7062	0.0403	0.1248	0.1514
AXL+CREB+LKB1+Src+STAT3	19	16	0.7625	0.0341	0.0874	0.0856
AXL+YAP1+CREB+LKB1+Src	19	16	0.6688	0.0838	0.2047	0.2392

También aparecen en color azul otros dos modelos de regresión ajustados con valores significativos, tanto a nivel de covariables, como de modelo global, este último al menos en el test de razón de verosimilitud (Likekihook).

En base a los resultados obtenidos en los diferentes modelos ajustados, parece ser que las variables *CREB*, *LKB1* y *Src,* muestran significancia en los diferentes modelos ajustados y parecen influir en el tiempo libre de progresión. Si se revisa el gráfico Gr.4.7 de los resultados del test de correlación de Spearman, vemos que algunos de los genes en los que se indicaba que existía correlación, son los que se determinan como significativos en el modelo de Cox ajustado.

A continuación se muestra el resumen completo del modelo (en rojo en la tabla), donde se pueden ver todos los estadísticos, entre otros los coeficientes β, los ratios $HR = exp(coef)$ (*estimación de riesgos relativos*), la desviación estándar y los intervalos de confianza de cada variable.

```
Formula =  ObjSLP ~ ddCt.CREB + ddCt.LKB1 + ddCt.Src + ddCt.STAT3

               coef   HR=exp(coef)   p-value  lower.95  upper.95
ddCt.CREB    0.5208         1.6834    0.0146    1.1081    2.5574
ddCt.LKB1   -7.2842         0.0007    0.0059    0.0000    0.1231
ddCt.Src     0.1629         1.1769    0.0464    1.0026    1.3816
ddCt.STAT3  -0.7286         0.4826    0.2026    0.1573    1.4803

Concordance = 0.756  (se = 0.092 )

Rsquare = 0.454   (max possible= 0.98 )
Likelihood ratio test = 11.49  on 4 df,   p=0.02
Wald test             =  9.56  on 4 df,   p=0.05
Score (logrank) test  =  9.27  on 4 df,   p=0.05
```

Interpretando el modelo resultante, a partir de los coeficientes β y la estimación de riesgo HR de cada covariable, se puede concluir con significancia de 0.05,

que el aumento de un punto de expresión del gen *CREB* incrementa un 68% $(HR = 1.68)$ el riesgo de muerte de los pacientes. En el caso del gen *Scr*, a igualdad de valores de expresión de los otros genes, el aumento de un punto en su expresión, significa un aumento del 18% $(HR = 1.18)$ en el riesgo de muerte, mientras que en el caso de *LKB1* el aumento de su nivel de expresión, manteniendo la expresión de los otros genes, tendría un efecto de disminución del riesgo de muerte ya que vemos que el valor de su coeficiente $\beta = -7.28$ es negativo, pero dado que su valor de $HR = 0.0007$ es muy cercano a cero, este no tendría una afectación significante en el pronóstico. Por otra parte, se puede comprobar que el intervalo de confianza de la covariable *HEST1* contiene el *1*, es decir contiene el valor nulo del modelo, por lo que no tiene una contribución significativa en el modelo.

Se ha efectuado un proceso de evaluación de validez del modelo mejor ajustado, con el test de riesgos proporcionales que ha indicado que no es significativo al 5% ni globalmente, ni para las diferentes covariables del modelo. Se puede concluir que no se viola el supuesto de riesgos proporcionales. También se ha efectuado el test individual de Schoenfeld para las betas resultantes del modelo.

Todo el detalle de los modelos de ajuste realizados, los diferentes test de validación y gráficas generadas, se pueden consultar en el informe de análisis completo (anexo II, apartado 8.2).

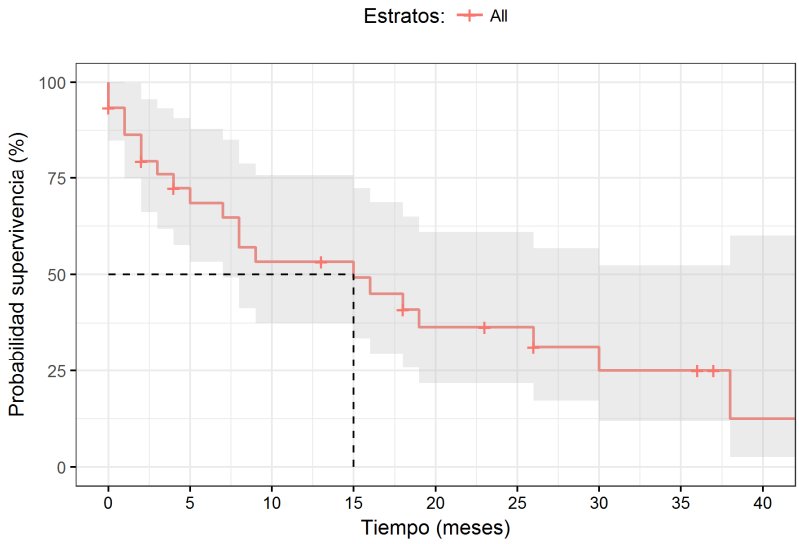
4.2.3.2.2 Supervivencia Global (SGL)

Para calcular la variable *time* para el análisis de Supervivencia Global (SGL), al igual que para el proceso de análisis SLP, como fecha de inicio del estudio se han tenido en cuenta las mismas fechas con los condicionantes de disponibilidad descritos en el apartado 4.2.3.2. El tiempo se ha calculado también en meses. Como fecha de producción del evento, se tomó la fecha de éxitus del paciente y en caso de desconocerse ésta se tomó la última fecha de seguimiento disponible como fin de periodo, marcando estos pacientes como censurados (en la tabla 4.2, la variable *ObjSGL* de los pacientes censurados tiene un símbolo +).

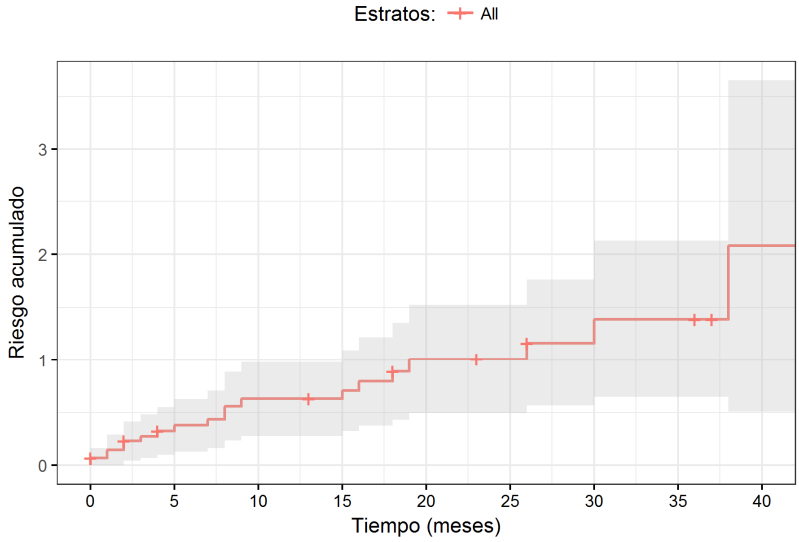
Para realizar el análisis SGL se han seguido las mismas estrategias que en el análisis SLP: SGL sin y con estratificación mediante estimación de *Kaplan-Meier* y modelos de regresión de Cox Univariable y Mutivariable. A continuación se indican únicamente los resultados que han presentado evidencia estadística significativa de que alguna de las expresiones génicas analizadas son factores influyentes en la supervivencia global (SGL) de los pacientes. En el informe de análisis completo se pueden consultar los detalles de cada análisis realizado.

4.2.3.2.2.1 Análisis de la SLP global sin estratificación

Al realizar el modelo de ajuste de supervivencia mediante la estimación de Kaplan-Meier, la mediana de Supervivencia Global (50% probabilidad de supervivencia) obtenida ha sido de 15 meses y la media de 18.20 meses. En las gráficas Gr.4.14 se muestra la curva resultante del modelo global y en la gráfica Gr.4.15 la curva de residuos acumulados.



Gráfica Gr. 4.14. *Curva de Supervivencia GL*



Gráfica Gr. 4.15. *Riesgo acumulado SGL*

4.2.3.2.2.2 Análisis de la GSL global estratificado por género y tipo de mutación KRAS

Al igual que en el análisis SLP, no se han encontrado evidencias significativas de diferencias entre los estratos analizados. Los detalles del resultado numérico y gráficos de ambos análisis se pueden consultar en el informe completo anexo (anexo II, apartado 8.2).

4.2.3.2.2.3 Ajuste de modelos de regresión de Cox Univariable para SGL

En la tabla 4.11, se muestra el resultado de la ejecución de la función para el modelo de ajuste de regresión de Cox univariable sin estratificación para la variable *ObjSGL* con cada covariable de expresión de cada gen.

Tabla 4.11. *Resultados del modelo de regresión de Cox Univariable (GL)*

	n	n.eventos	beta	HR (95% CI for HR)	wald.test	p.value	CutP
ddCt.HES1	19	13	-0.1284	0.8795 (0.2912-2.6563)	0.05	0.82	0.34
ddCt.CDCP1	28	18	-0.0043	0.9957 (0.9581-1.0347)	0.05	0.8256	15.9
ddCt.AXL	25	15	0.0735	1.0763 (0.7881-1.4698)	0.21	0.6439	3.69
ddCt.YAP1	25	16	-0.0187	0.9815 (0.8682-1.1094)	0.09	0.7648	1.77
ddCt.CREB	23	14	-0.3565	0.7001 (0.4662-1.0514)	2.95	0.0857	4.45
ddCt.LKB1	20	12	-5.4628	0.0042 (0-0.5831)	4.73	0.0297	0.43
ddCt.Src	28	19	0.0082	1.0082 (0.9898-1.0269)	0.75	0.3857	5.07
ddCt.STAT3	26	17	0.3364	1.3999 (0.4401-4.4528)	0.32	0.5689	0.37

Se puede apreciar que la expresión del gen *LKB1* es significativa al 5% con $p - valor = 0.0297$, con un $HR = 0.0042$ y un intervalo de confianza de $(0 - 0.5831)$, que nos indica que la expresión del gen influye en la supervivencia global de los pacientes.

Estratificación según expresión Alta, Media y Baja

En rojo, en la tabla 4.12, la expresión del gen *LKB1* muestra evidencia de significación al *5%* en el estrato de expresión *Baja* respecto a la expresión *Alta*, con un $p - valor = 0.015$ y con un test de Wald también significativo con un $p - valor = 0.0397$.

Tabla 4.12. *Modelo de Cox Univariable Estratificado (Alta, Media y Baja) SGL*

Variable	n	n.eventos	wald.test	p.value	strat	beta	HR (95% CI for HR)	Pr(>|z|)
HES1.st	19	13	0.78	0.6774	HES1.stBaja	0.6102	1.8408 (0.452-7.4963)	0.3944
					HES1.stMedia	0.4326	1.5413 (0.4121-5.7643)	0.5204
CDCP1.st	28	18	1.9	0.3863	CDCP1.stBaja	0.4717	1.6027 (0.5027-5.1102)	0.4253
					CDCP1.stMedia	-0.3374	0.7136 (0.1996-2.5514)	0.6038
AXL.st	25	15	0.72	0.6974	AXL.stBaja	-0.4593	0.6317 (0.1608-2.4825)	0.5107
					AXL.stMedia	0.0678	1.0702 (0.3107-3.6858)	0.9144
YAP1.st		16	0.17	0.9188	YAP1.stBaja	-0.1574	0.8543 (0.2386-3.0595)	0.8089
					YAP1.stMedia	0.1307	1.1396 (0.3354-3.8723)	0.8342
CREB.st	23	14	2.26	0.3224	CREB.stBaja	1.0689	2.9121 (0.7198-11.7817)	0.1339
					CREB.stMedia	0.7669	2.1531 (0.5074-9.1362)	0.2984
LKB1.st	20	12	6.45	0.0397	LKB1.stBaja	2.6439	14.0681 (1.6694-118.5546)	0.015
					LKB1.stMedia	1.7932	6.0084 (0.7002-51.5612)	0.1021
Src.st	28	19	0.88	0.6439	Src.stBaja	-0.398	0.6717 (0.2014-2.2398)	0.5172
					Src.stMedia	0.1358	1.1455 (0.3922-3.3452)	0.8038
STAT3.st	26	17	0.43	0.8075	STAT3.stBaja	-0.2844	0.7525 (0.2075-2.7286)	0.6652
					STAT3.stMedia	0.143	1.1537 (0.3713-3.585)	0.8048

Estratificación según expresión Alta, Baja-Media

Con esta estratificación, tal como reflejan los resultados de los diferentes modelos de Cox univariables (tabla 4.13), también existe evidencia significativa al *5%* en el gen *LKB1*, con un $p - valor = 0.0389$ entre expresión *Alta* versus *Baja-Media*.

Tabla 4.13. *Modelo de Cox Univariable Estratificado (Alta, Baja-Media) SGL*

Variable	n	n.eventos	wald.test	p.value	strat	beta	HR (95% CI for HR)
HES1.st	19	13	0.7	0.4024	HES1.stBaja-Media	0.5061	1.6589 (0.5073-5.4244)
CDCP1.st	28	18	0.02	0.8932	CDCP1.stBaja-Media	0.0727	1.0754 (0.3723-3.1064)
AXL.st	25	15	0.09	0.7641	AXL.stBaja-Media	-0.1732	0.841 (0.2713-2.6065)
YAP1.st	25	16	0	0.9823	YAP1.stBaja-Media	-0.0118	0.9883 (0.349-2.7984)
CREB.st	23	14	1.99	0.1582	CREB.stBaja-Media	0.922	2.5143 (0.6986-9.0486)
LKB1.st	20	12	4.27	0.0389	LKB1.stBaja-Media	2.1602	8.6731 (1.1165-67.3727)
Src.st	28	19	0.04	0.8399	Src.stBaja-Media	-0.1013	0.9037 (0.3383-2.4143)
STAT3.st	26	17	0.01	0.9336	STAT3.stBaja-Media	-0.0434	0.9575 (0.3448-2.6589)

Estratificación según expresión según expresión <Q25 y >=Q25

Tabla 4.14. *Modelo de Cox Univariable Estratificado <Q25 y >=Q25 SGL*

Variable	n	n.eventos	wald.test	p.value	strat	beta	HR (95% CI for HR)
HES1.st	19	13	4.45	0.0349	HES1.st>=Q25	-1.5201	0.2187 (0.0533-0.8979)
CDCP1.st	28	18	3.95	0.0469	CDCP1.st>=Q25	-0.9896	0.3717 (0.1401-0.9864)
AXL.st	25	15	0.23	0.6315	AXL.st>=Q25	0.2892	1.3354 (0.4095-4.3544)
YAP1.st	25	16	0.17	0.6824	YAP1.st>=Q25	0.2681	1.3074 (0.3621-4.7211)
CREB.st	23	14	0.26	0.6132	CREB.st>=Q25	-0.3093	0.7339 (0.2212-2.4353)
LKB1.st	20	12	8.44	0.0037	LKB1.st>=Q25	-2.1632	0.115 (0.0267-0.4949)
Src.st	28	19	1.36	0.2432	Src.st>=Q25	0.7404	2.0968 (0.6047-7.2712)
STAT3.st	26	17	2.17	0.1404	STAT3.st>=Q25	1.1321	3.1021 (0.6887-13.9731)

El resultado obtenido con esta estratificación se puede ver en la tabla 4.14. Se muestran tres covariables que tienen evidencia de significación al 5% (en color rojo), *HES1*, *CDCP11* y *LKB1*, en su estrato de $>= Q25$ respecto a $< Q25$, con $p - valores\ de\ 0.0349,\ 0.0469\ y\ 0.0037$, respectivamente. Los $HR\ son\ de\ 0.22,\ 0.37\ y\ 0.12$.

Estratificación según expresión con mejores puntos de corte

Los mejores puntos de corte para el modelo de Cox de SGL se muestran en la tabla 4.15.

Tabla 4.15 *Estratificación a partir de mejores puntós de corte*

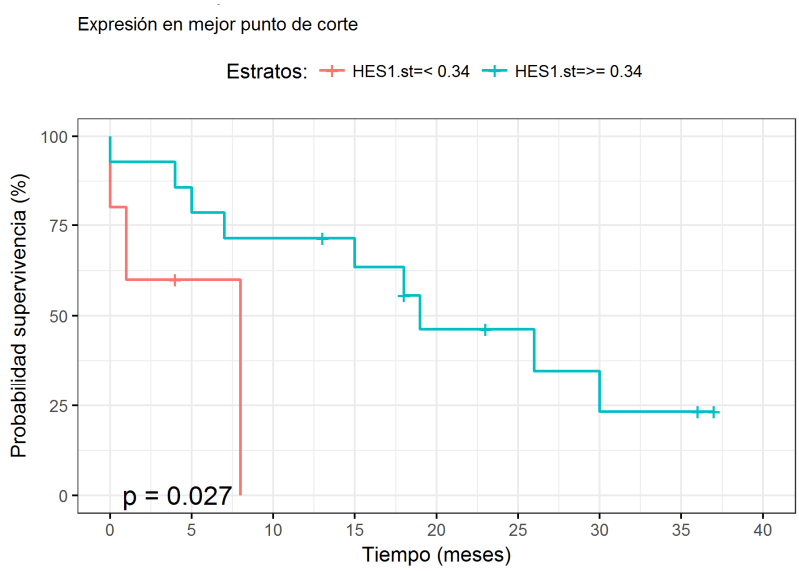
HES1.st	CDCP1.st	AXL.st	YAP1.st	CREB.st	LKB1.st	Src.st	STAT3.st
< 0.34 : 5	< 15.9 :13	< 3.69 : 7	< 1.77 : 2	< 4.45 :20	< 0.43 : 7	< 5.07 :16	< 0.37 : 7
>= 0.34 :14	>= 15.9 :15	>= 3.69 :18	>= 1.77 :23	>= 4.45 : 3	>= 0.43 :13	>= 5.07 :12	>= 0.37 :19
NA's :11	NA's : 2	NA's : 5	NA's : 5	NA's : 7	NA's :10	NA's : 2	NA's : 4

En la tabla 4.16 se muestran los resultados de los diferentes modelos de Cox ajustados para cada gen, aplicando la estratificación del mejor punto de corte. En rojo se indican los modelos que han resultado significativos al 5% y podemos ver que son los genes *HES1* y *LKB1*, con un $p - valor\ de\ 0.0349\ y\ 0.0135$ y un $HR\ de\ 0.22\ y\ 0.23$, respectivamente.

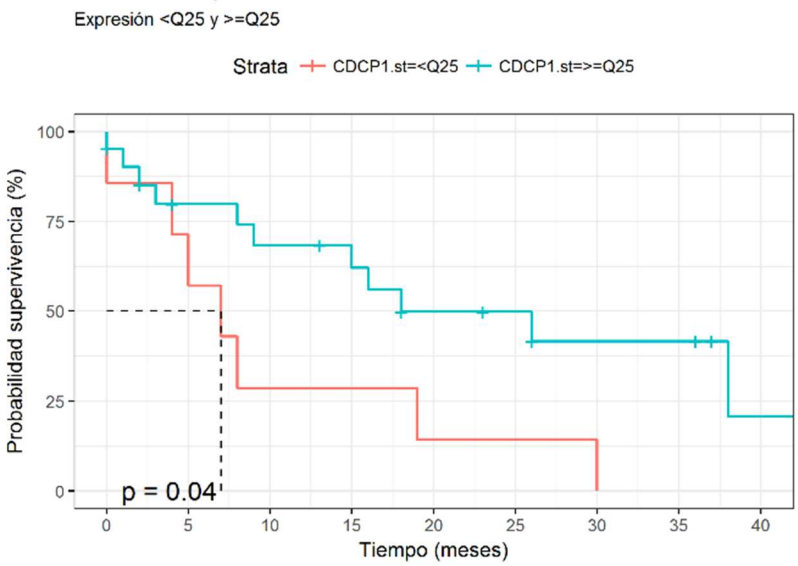
***Tabla 4.16**. Modelo de Cox Univariable estratificado con mejores puntos de corte*

Variable	n	n.eventos	wald.test	p.value	strat	beta	HR (95% CI for HR)
HES1.st	19	13	4.45	0.0349	HES1.st>= 0.34	-1.5201	0.2187 (0.0533-0.8979)
CDCP1.st	28	18	1.93	0.1652	CDCP1.st>= 15.9	-0.6894	0.5019 (0.1896-1.3288)
AXL.st	25	15	0.53	0.466	AXL.st>= 3.69	0.4345	1.5442 (0.4801-4.9666)
YAP1.st	25	16	0	0.9982	YAP1.st>= 1.77	19.5172	299361866.8478 (0-Inf)
CREB.st	23	14	0	0.9981	CREB.st>= 4.45	-19.5095	0 (0-Inf)
LKB1.st	20	12	6.1	0.0135	LKB1.st>= 0.43	-1.4624	0.2317 (0.0726-0.7396)
Src.st	28	19	1.12	0.2907	Src.st>= 5.07	0.5044	1.656 (0.6497-4.221)
STAT3.st	26	17	2.17	0.1404	STAT3.st>= 0.37	1.1321	3.1021 (0.6887-13.9731)

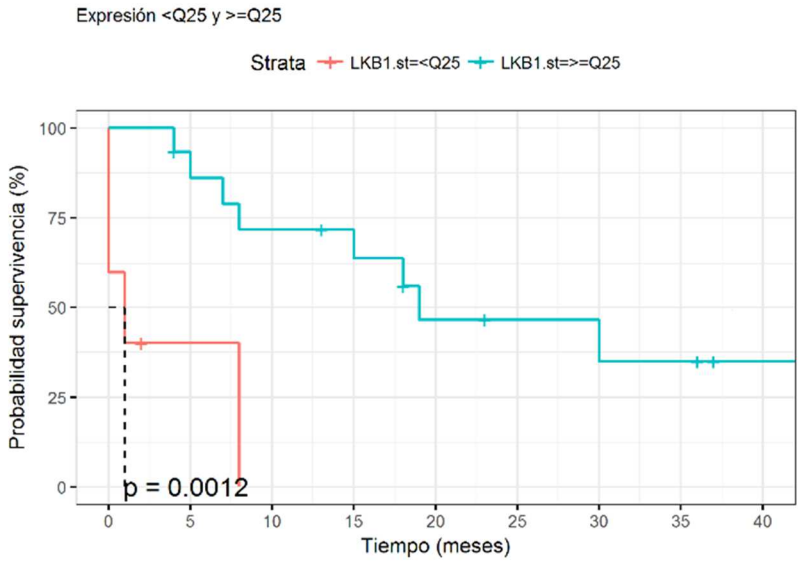
En las gráficas Gr.4.16, Gr.4.17 y Gr.4.18 se muestran las curvas de supervivencia de los tres modelos que han tenido evidencia significativa al *5%* en alguna de las estratificaciones.



***Gráfica Gr. 4.16.** Curva de Supervivencia HES1 estratificado*



***Gráfica Gr. 4.17.** Curva de Supervivencia CDCP1 estratificado*



Gráfica Gr. 4.18. *Curva de Supervivencia LKB1 estratificado*

Todos los valores de expresión que han resultado significativos se han validado con un test de Log-Rank de comparación por pares, cuyo resultado confirma la significación de todos ellos.

4.2.3.2.2.4 Ajuste de modelos de regresión de Cox Multivariable para SLP

En las tablas 4.17 y 4.18 que se muestran a continuación, vemos que de todos los modelos ajustados solo hay tres modelos que tienen significancia global al *5%* según el test de ajuste de razón de verosimilitud (Likekihook), pero solo uno (línea en rojo) es significativo adicionalmente en el test de Wadl y en el test de puntajes o score. Por tanto, parece que el más ajustado es el que utiliza las dos covariables de la expresión del gen *LKB1* y *Scr*, las cuales tienen significancia al *5%*. El índice de concordancia del modelo $C = 0.79$, es bastante correcto. Solo se han podido utilizar 20 muestras de las 30 de conjunto de datos inicial y se han producido 12 eventos.

En conclusión, se puede ver que en el caso de SGL el resultado obtenido con el modelo de Cox Univariable solo presenta significancia en el gen *LKB1* y el modelo Cox Multivariable, presenta significancia en las covariables *LKB1* y *Src*.

Tabla 4.17. *Variables significativas de cada modelo de regresión de Cox SGL*

Modelo	Var.Significativas	p.valores
HES1+CDCP1+AXL+YAP1+CREB+LKB1+Src+STAT3	LKB1	0.0276
CDCP1+AXL+YAP1+CREB+LKB1+Src+STAT3	LKB1, Src	0.0313, 0.0451
CDCP1+YAP1+CREB+LKB1+Src+STAT3	LKB1	0.0277
AXL+YAP1+CREB+LKB1+Src+STAT3	LKB1	0.0312
YAP1+CREB+LKB1+Src+STAT3	LKB1	0.0328
CREB+LKB1+Src+STAT3	LKB1	0.0467
CREB+LKB1+Src	LKB1	0.0478
LKB1+Src+STAT3	LKB1, Src	0.0059, 0.0456
LKB1+Src	LKB1, Src	0.0058, 0.0494

Tabla 4.18. *Significación global de cada modelo*

Modelo	n	n.eventos	Conc.	likelih.pval	wald.pval	score.paval
HES1+CDCP1+AXL+YAP1+CREB+LKB1+Src+STAT3	17	11	0.781	0.2243	0.5867	0.3302
CDCP1+AXL+YAP1+CREB+LKB1+Src+STAT3	19	11	0.7881	0.103	0.5591	0.4758
CDCP1+YAP1+CREB+LKB1+Src+STAT3	19	11	0.7627	0.0859	0.4194	0.3759
AXL+YAP1+CREB+LKB1+Src+STAT3	19	11	0.7712	0.1224	0.3664	0.4212
YAP1+CREB+LKB1+Src+STAT3	19	11	0.7712	0.0779	0.2779	0.306
CREB+LKB1+Src+STAT3	19	11	0.7881	0.063	0.2174	0.2625
CREB+LKB1+Src	19	11	0.7203	0.0329	0.1335	0.1647
LKB1+Src+STAT3	20	12	0.8015	0.0061	0.0509	0.0624
LKB1+Src	20	12	0.7941	0.0022	0.0206	0.03

A continuación se detalla el modelo completo de ajuste indicado en rojo en la tabla, donde se pueden ver todos los estadísticos, entre otros los coeficientes *β*, los ratios *HR* (estimación de riesgos relativos), la desviación estándar y los intervalos de confianza de cada variable.

```
Formula =  ObjSLP ~ ddCt.LKB1 + ddCt.Src
                    coef    exp(coef)      p-value     lower.95     upper.95
ddCt.LKB1    -7.8008989   0.0004094    0.00582      0.0000       0.1046
ddCt.Src      0.1941104   1.2142303    0.04945      1.0005       1.4737

Concordance= 0.794  (se = 0.1 )
Rsquare= 0.458   (max possible= 0.946 )
Likelihood ratio test = 12.24  on 2 df,   p=0.002
Wald test             =  7.76  on 2 df,   p=0.02
Score (logrank) test  =  7.01  on 2 df,   p=0.03
```

Si se analiza la estimación de riesgo $HR = exp(coef)$ en el resumen mostrado de cada covariable, se puede concluir con significancia de 0.05 que el modelo indica que a igualdad de valores en el resto de expresiones de los otros genes, el aumento de un punto de expresión del gen *Scr*, significa un aumento del 21.4% $(HR = 1.214)$ en el riesgo de muerte de los pacientes, mientras que en el caso de *LKB1* el aumento de su nivel de expresión, manteniendo la expresión de los otros genes, tendría un efecto de disminución del riesgo de muerte ya que vemos que el valor de su coeficiente $\beta = -7.80$ es negativo, pero dado que su valor de $HR = 0.0004$ es muy cercano a cero, este no tendría una afectación significante en el pronóstico.

Al igual que con el modelo de regresión de Cox de SLP, se ha efectuado el test de riesgos proporcionales que ha indicado que no es significativo al 5% ni globalmente, ni para las diferentes covariables del modelo. Se puede concluir que no se viola el supuesto de riesgos proporcionales. También se ha efectuado el test individual de Schoenfeld para las betas resultantes del modelo.

Todo el detalle de los modelos de ajuste realizados, los diferentes test de validación y gráficas generadas, se pueden consultar en el informe de análisis completo (anexo II, apartado 8.2).

4.2.4 Conclusiones del análisis de supervivencia

Como primer aspecto a destacar del proceso de análisis bioestadístico realizado, que surge tras la preparación y exploración de los datos obtenidos del estudio real, es el número reducido de muestras finales disponibles (30) para el estudio, que además no son de forma homogénea completas para cada muestra, con lo que ello comporta para los diferentes análisis bioestadísticos realizados.

El análisis de supervivencia está enfocado a dilucidar si existe evidencia significativa sobre si la expresión de los genes analizados molecularmente son influyentes en la Supervivencia Libre de progresión (SLP) o en la Supervivencia General (SGL) de los pacientes, teniendo en cuenta el condicionante de que no siempre se dispone de la información de expresión de todos los genes. Aunque en los procesos de análisis reales este factor es siempre complejo, hay que considerar que los resultados obtenidos serían más precisos si el número de muestras fuera algo mayor y/o más homogéneo en cuanto a los datos obtenidos de cada paciente.

Por otra parte, también hay que mencionar que a diferencia de lo que sería un estudio teórico convencional de Supervivencia, en los casos de análisis de supervivencia relacionados con la progresión del cáncer, los periodos de análisis siempre estarán basados en los periodos entre las fechas en los que los pacientes fueron tratados, que evidentemente pueden variar entre las muestras, no siendo entre fechas fijas determinadas en el tiempo. Por ello, el periodo de análisis establecido para este estudio real de pacientes, no se ha fijado en un espacio temporal de inicio concreto, sino que se marca como inicio del tiempo de análisis la fecha de inicio del tratamiento de primera línea que se administra al paciente, por lo que abarca diferentes años naturales. Este aspecto ha generado una dificultad añadida en el proceso del cálculo de los tiempos (time) de supervivencia, dado que la información sobre las características de las diferentes variables (fechas clave, progresión de la enfermedad, etc.) presentaba algunas dudas que se han tenido que resolver directamente del responsable del estudio original.

En la muestra analizada, el *73.33%* de los pacientes eran hombres y el *26.67%* eran mujeres, dato que como se ha podido comprobar difiere de las estadísticas poblacionales analizadas sobre este tipo de cáncer en el capítulo 3. Posiblemente esta diferencia está relacionada con la *"n"* reducida de la muestra tratada y solo referente a pacientes con la mutación en *KRAS*. Los hombres muestran una media de consumo de tabaco el doble superior al de las mujeres (Gr. 4.3). No existe mucha diferencia de edades entre sexos, la media de edad de los pacientes está en el rango de entre 54 y 60 años, mujeres y hombres, respectivamente (Gr. 4.1).

En referencia a la histología de los cánceres muestreados, el porcentaje mayor es de Adenocarcinomas con un 80% y las mutaciones más habituales de KRAS han sido las G12C, la G12V y la G12D (Gr. 4.4 y Gr. 4.5).

Se ha detectado la existencia de correlación significativa (del *0.05* o inferior) entre la expresión de varios genes. Los genes *LKB1* y *CREB* son los que presentan una mayor correlación en su expresión, seguida de los genes *YAP1* y *SRC*, de *YAP1* y *STAT3* y con menor fuerza de correlación *STAT3* y *HES1* (gráfica Gr. 4.7). Esta correlación está relacionada con la significación estadística encontrada en estos genes como factor de influencia en la SLP y SLG.

En referencia a la estimación de supervivencia (SLP) mediante el modelo de Kaplan-Meier, se han tratado 30 muestras con 27 eventos y 3 censuras. La mediana de SLP general ha sido de 4.5 meses y la media de 8.2 meses (Gr. 4.8). En la estimación de SLP general estratificada por sexo y tipo de mutación de *KRAS*, no se han encontrado evidencias de diferencias significativas entre estos estratos.

Se han realizado diferentes análisis de supervivencia mediante el modelo de regresión de Cox. En el modelo univariable sin estratificar, no se han encontrado evidencia significativa de influencia de ninguno de los genes en la supervivencia de los pacientes. En el modelo univariable estratificado, se han encontrado evidencias significativas estadísticas al *5%* de que la expresión de los genes *LKB1* y *HES1* son factor influyente en el tiempo de supervivencia, estratificando su expresión en percentil de 25. En el caso de *LKB1* también se ha encontrado significación al *5%* entre la expresión de la estratificación Alta vs Baja. Como se puede apreciar en las gráficas Gr. 4.10, Gr. 4.11 y Gr.4.12, la baja expresión de los genes *LKB1* y *HES1* indican una supervivencia inferior de los pacientes.

Adicionalmente, se han analizado los mejores puntos de corte para todos los genes (tabla 4.7) mediante la función *cutp()* del paquetes *survMisc*, que implementa el modelo de obtención de Contal C y O'Quigley J (1999). A partir de los mejores puntos de corte se ha realizado una nueva estratificación de la expresión de los genes y se ha efectuado nuevos modelos de regresión de Cox con esta estratificación, que han denotado evidencia significativa al *5%* en los genes *HES1*, *LKB1* y *Src* (tabla 4.8). En las gráficas Gr.13a-c se muestran las curvas de supervivencia de los genes significativos, obtenidas con los mejores puntos de corte, donde se puede apreciar los límites en la expresión de estos genes y su influencia en la supervivencia de los pacientes. Se puede ver que en el caso de la expresión gen *Src*, a diferencia de la de los genes *HES1* y *LKB1*, los niveles altos de expresión tienen un pronóstico peor para la supervivencia de los pacientes.

Asumiendo la linealidad de las variables (difícil de validar dado en bajo número de muestras) y verificando la suposición de riesgos proporcionales, se han realizado varios procesos de ajuste con diferentes modelos de regresión de Cox multivariable. Varios de estos modelos han evidenciado significancia al *5%* de varios genes, siendo el más destacado el modelo con los genes *CREB* + *LKB1* + *Scr* + *STAT3*, donde *CREB*, *LKB1* y *Src* indicaban evidencia *estadística* significativa al *5%* y con un índice de Concordancia del modelo $C = 0.7562$, indicando con ello la influencia de estos genes en el tiempo de SLP. Interpretando el modelo, nos indica que con significancia del *5%* y a igualdad de valores en el resto de expresiones de los otros genes, el aumento de un punto de expresión del gen *CREB* incrementa un 68% $(HR = 1.68)$ el riesgo de muerte. En el caso del gen *Scr*, a igualdad de valores de expresión de los otros genes, el aumento de un punto en su expresión, significa un aumento del 18% $(HR = 1.18)$ en el riesgo de muerte, mientras que en el caso de *LKB1* el aumento de su nivel de expresión, manteniendo la expresión de los otros genes, tendría un efecto de disminución del riesgo de muerte ya que vemos que el valor de su coeficiente $\beta = -7.28$ es negativo, pero dado que su valor de $HR = 0.0007$ es muy cercano a cero, este no tendría una afectación significante en el pronóstico. Adicionalmente, se puede comprobar que el intervalo de confianza de la covariable *HEST1* contiene el *1*, es decir contiene el valor nulo del modelo, no tiene una contribución significativa en el modelo.

En el apartado de análisis de SLP se pueden ver en los diferentes gráficos y tablas que aportan los datos numéricos y gráficos de los resultados obtenidos. Adicionalmente, se han analizado las posibles interacciones entre las covariables, determinado que no hay evidencia significativa de ninguna

interacción entre ellas. Se han efectuado las pruebas de riesgos proporcionales, dando esta que no son significativas.

En referencia a la estimación de supervivencia global (SGL) mediante el modelo de Kaplan-Meier, se han tratado 30 muestras con 20 eventos y 10 censuras. La mediana de SGL general ha sido de 15 meses y la media de 18.2 meses (Gr. 4.14). En la estimación de SGL general estratificado por sexo y tipo de mutación de *KRAS*, no se han encontrado evidencias de diferencias significativas para estos estratos.

Al igual que en la SLP, para el análisis de la SGL se han realizado diferentes análisis de ajuste de las covariables, mediante diferentes modelos de regresión de Cox Univariable. Al efectuar el modelo de ajuste Univariable no estratificado, se aprecia evidencia significativa al *5%* solo del gen *LKB1* de influencia en el tiempo de supervivencia global de los pacientes. En el modelo Univariable estratificado, hay evidencia significativa al *5%* tanto en la estratificación de *Alta* versus *Baja*, como en la estratificación *Baja-Media versus Alta*, también en el gen *LKB1*. En la estratificación por cuantiles en el percentil >=25 o <25, aparecen tres genes significativos al *5%*, que han sido el *HES1*, *CDCP1* y *LKB1* con $p-valores\ de\ 0.0349, 0.0469\ y\ 0.0037$, respectivamente. Analizando los mejores puntos de corte para el nivel de expresión con la función *cutp()* (tabla 4.15), se han efectuado nuevos modelos de regresión de Cox con esta estratificación, que han denotado evidencia significativa al 5% en los genes *HES1* y *LKB1* (tabla 4.17). En las gráficas Gr. 4.16, Gr. 4.17 y Gr. 4.18 se muestran las curvas de supervivencia de los genes significativos según los mejores puntos de corte, donde se puede apreciar los límites en la expresión de estos genes y su influencia en la supervivencia de los pacientes.

Asumiendo igualmente la linealidad de las covariables, se han efectuado varios modelos de regresión de Cox con las diferentes covariables de entrada, donde solo se ha encontrado significancia al *5%* a nivel de test de razón de verosimilitud (Likekihook) en dos de los modelos ajustados: en el modelo con las variables de entrada *LKB1* + *Src* + *STAT3*, con significancia solo las covariables *LKB1* y *Src*; en el modelo con las covariables de entrada *LKB1* + *Src*, además de dar significancia el test global de razón de verosimilitud (Likekihook), también han sido significativos al 5% los test de Wald y de score. Interpretando el modelo lineal, podemos entender que el incremento de un punto en la expresión de *Src*, representa un 21.4% de incremento en el riesgo de mortalidad de los pacientes. En el caso de *LKB1* el aumento de su nivel de expresión, manteniendo la expresión de los otros genes, tendría un efecto de disminución del riesgo de muerte ya que vemos que el valor de su coeficiente $\beta = -7.80$ es negativo, pero dado que su valor de $HR = 0.0004$ es muy cercano a cero, este no tendría una afectación significante en el pronóstico.

Se han efectuado las pruebas de riesgos proporcionales, dando estas que no son significativas.

Después de ver los resultados de los diferentes análisis de supervivencia realizados, se puede concluir que la expresión de los genes que presentan evidencia significativa al *5%* y que por tanto tienen influencia en el tiempo de supervivencia de los pacientes con cáncer de pulmón (NSCLC) con mutación en *KRAS* ha sido la relativa a los genes *LKB1*, *Src*, *HES1*, *CREB* y *CDCP1*.

4.2.5 Comparativa de resultados con estudio original

A nivel global se han encontrado algunas diferencias entre los resultados obtenidos en el estudio original y los presentados en este análisis. Como se indicaba en la descripción del estudio original, la reducida cantidad de muestras y el no disponer de forma homogéneamente completa los diferentes datos para todas las muestras, puede haber sido un factor fundamental para que puedan existir algunas discrepancias en los cálculos realizados y por tanto en los resultados obtenidos. El cálculo del tiempo de supervivencia de cada muestra presentó dificultades por la falta de información clara respecto a los diferentes límites de cálculo, tanto para la fecha de inicio de análisis, como para la fecha de progresión (este último debido a la falta de conocimientos médicos sobre este aspecto), que finalmente se pudieron solventar directamente con el responsable del trabajo original.

A continuación se efectúa una revisión de los diferentes resultados obtenidos en ambos trabajos, indicando las divergencias y concordancias encontradas entre ambos análisis.

- En ambos análisis se ha trabajado con 30 muestras, de las cuales se han identificado 27 eventos y 3 censuras para la SLP y 20 eventos y 10 censuras para la SGL. La exploración y revisión de los datos disponibles para el estudio han mostrado una total convergencia en los resultados mostrados en ambos trabajos.
- En referencia al análisis de correlación entre las diferentes expresiones de los genes, han mostrado alguna divergencia de parejas de correlación, aun aplicando el test e Spearman en ambos casos:
  - En el estudio original se ha detectado correlación significativa al 5% en 7 parejas, mientras en el presente estudio solo han mostrado correlación significativa en 4 de las 7 parejas del estudio original. Las parejas de *CREB1* y *Src*, *CREB1* y *STAT3*, así como de *Src* y *STAT3*, en el presente análisis (gráfica Gr. 4.7) no presentaban correlación significante al 5%. La discrepancia en este aspecto, podría venir causada porque en el estudio original se hubieran empleado 2 muestras que en este trabajo se han descartado del estudio, dada la inexistencia de las fechas de análisis necesarias para realizar el estudio de supervivencia, aunque estas dos muestras tenían los valores de expresión de los diferentes genes. En este estudio se han descartado estas dos muestras para todos los análisis realizados. Se desconoce si en el estudio original se han tenido en cuenta para algunos de los análisis realizados en él.
- En el análisis de Supervivencia Libre de Progresión (SLP) global existe discrepancia en la mediana del tiempo de progresión:
  - En el estudio original se indica que la mediana de SLP es de 2.5 meses y no se describe la media.
  - En el presente estudio se confirma que la mediana de SLP es de 4.5 meses y la media es de 8.19 meses.
- Del análisis de supervivencia SLP estratificado por género y por tipo de mutación de KRAS no se puede hacer comparativa, dado que es un análisis

no realizado en el estudio original. En el presente trabajo se ha podido constatar que no hay evidencia significativa de diferencias entre la supervivencia entre hombres y mujeres, ni en base a las diferentes mutaciones del gen KRAS.

- En el análisis mediante un modelo de regresión de Cox Univariable sin estratos, se obtiene el mismo resultado en ambos trabajos, es decir no hay evidencia significativa de la influencia de la expresión de los genes en el tiempo de supervivencia.
- En el análisis mediante modelos de regresión de Cox Univariables estratificados, se han realizado las mismas agrupaciones de expresión en los genes, aunque los estadísticos obtenidos no eran iguales en ambos trabajos. La comparativa de los resultados obtenidos es la siguiente:
  - En la estratificación entre niveles de expresión *Alta*, *Media* y *Baja*, en el estudio original no se ha encontrado evidencia significativa de la influencia de la expresión de ninguno de los genes analizados en el tiempo de supervivencia, en cambio, en el presente trabajo se han encontrados evidencia significativa al 5% en la expresión *Alta* versus *Baja* del gen *LKB1*.
  - En la estratificación entre el nivel de expresión *Alta* y *Baja-Media*, es al contrario, en el estudio original se encuentra evidencia significativa al 5% en el gen Src, mientras que en el presente trabajo no se encuentra en ninguno de los genes analizados con esta agrupación.
  - En la estratificación según percentiles $>= 25$, en el estudio original se encuentra evidencia significativa al 5% entre la expresión $>= Q25$ vs $< Q25$ en el gen *LKB1*. En el presente trabajo se encuentra adicionalmente al gen *LKB1*, también significancia en el gen *HES1*.
  - En ambos estudios se ha realizado el cálculo de los mejores puntos de corte para el modelo de Cox Univariable de cada expresión génica y ha habido coincidencia en dos de los genes que presentaban evidencia significativa al 5% en la estratificación con el mejor punto de corte, que ha sido en los genes *LKB1* y *Src*, pudiendo verificar en estos casos que los puntos de corte en ambos casos eran los mismos, pero se ha mostrado discrepancia en el resto de resultados. En el estudio original se han identificado los genes *CDCP1* y *CREB1* con significación estadística al 5% y en cambio en el presente estudio, solo en el gen *HES1*.
- En referencia al análisis con modelo de Cox Multivariable, no se puede realizar comparativa, ya que este análisis no se realizó en el estudio original.
- En el análisis de Supervivencia Global, también existe algo de discrepancia entre los valores de mediana obtenidos en ambos trabajos, aunque menor que la existente en la SLP:
  - En el trabajo original se obtuvo una mediana de 13.4 meses de SLP. La media no se indicaba.
  - En el presente trabajo se obtuvo una mediana de 15 meses y una media de 18.2 meses.

- Del análisis de supervivencia SGL estratificado por género y por tipo de mutación de KRAS no se puede hacer comparativa, dado que es un análisis no realizado en el estudio original. En el presente trabajo se ha podido constatar que tampoco hay evidencia significativa de diferencias entre la supervivencia entre hombres y mujeres, ni en base a las diferentes mutaciones del gen KRAS a nivel global.
- En el análisis mediante un modelo de regresión de Cox Univariable sin estratos, se obtiene el mismo resultado en ambos trabajos, ambos muestran evidencia significativa al 5% de la influencia de la expresión de los genes en el tiempo de supervivencia para el gen *LKB1*.
- En el análisis mediante modelos de regresión de Cox Univariables estratificados, se han realizado solo algunas agrupaciones de expresión equivalentes en los genes y la comparativa de los resultados obtenidos en las estratificaciones equivalentes es la siguiente:
  - En las estratificaciones entre niveles de expresión *Alta*, *Media* y *Baja* y entre *Alta*, *Baja-Media*, son equivalentes en ambos trabajos, se han encontrado evidencia significativa al 5% en la expresión *Alta* versus *Baja* del gen *LKB1*, aunque los valores de los estadísticos obtenidos no eran idénticos.
  - Al igual que en la SLP, se ha realizado el cálculo de los mejores puntos de corte para el modelo de Cox Univariable de cada expresión génica y ha habido coincidencia solo en el gen *LKB1*, que ambos estudios han mostrado significancia, pero en el estudio original se encontró significancia, también en la expresión del gen *CDCP1* y en cambio en el presente trabajo se encontró en la expresión del gen *HES1*.
  - En el estudio original no se realizó la estratificación según percentiles $>= 25$, pero en el presente trabajo se realizó obteniendo evidencia significativa al 5% de influencia de los genes *HES1*, *CDCP* y *LKB1* en el tiempo de supervivencia global.
- Como puede observarse, si revisamos las coincidencias en la expresión de los genes que en algún punto han sido significativos al 5% como factores influyentes en la supervivencia de los pacientes, solo aparecen los genes *LKB1*, *Src* y *CDCP*.

Capítulo 5

5. Conclusiones

Partiendo de las líneas estratégicas globales que se identificaron como la base de justificación y motivación del presente trabajo, se hace necesario agrupar los resultados obtenidos en tres grandes ámbitos, coincidiendo cada uno de estos ámbitos con cada uno de los objetivos globales que se marcaron: implicación de la mutación de KRAS en la oncogénesis del cáncer de pulmón, incidencia del cáncer de pulmón en la población, y un análisis bioestadístico de un estudio real sobre la supervivencia de pacientes en base a la expresión génica.

En referencia a la implicación de la mutación del gen KRAS en el cáncer de pulmón, se ha podido concluir que la familia de proteínas RAS está implicada en múltiples vías de señalización celular MAPK, entre ellas la MAPK/ERK, donde la mutación más frecuente en la familia de RAS se encuentra en el codón 12 de la isoforma de KRAS. A su vez, se ha podido constatar que la complejidad en el funcionamiento de las vías de señalización, junto con las mutaciones génicas, entre otras en KRAS, dificulta considerablemente la identificación de dianas terapéuticas efectivas en el cáncer de pulmón de células no pequeñas (NSCLC), siendo ello aún un reto para la comunidad científica.

En el segundo ámbito de investigación del trabajo, se podido ver que el cáncer de pulmón tiene en nuestros días, por desgracia, una posición muy destacada en la mortalidad de cáncer en la población, tanto a nivel mundial como a nivel europeo. En Europa, el 20.1% de muertes de cáncer son por cáncer de pulmón y de estas muertes el 55.62% son hombres, respecto al 44.38% de mujeres. KRAS es el segundo gen mutado (19.5%) en el cáncer de pulmón relacionado con la vía MAPK/ERK.

Como resultado de los análisis de Supervivencia realizados en trabajo, sobre pacientes con cáncer de pulmón (NSCLC) que presentaban mutación de KRAS, se ha obtenido evidencia estadística significativa al 5% de la influencia de la expresión de 5 genes implicados (*LKB1*, *Src*, *HES1*, *CREB* y *CDCP1*) en la Supervivencia libre de progresión y/o global de los pacientes. Este hecho confirma la necesidad de analizar y estudiar los niveles de expresión de los genes implicados en las vías de señalización, para comprender mejor su funcionamiento y así identificar dianas terapéuticas más efectivas.

Dada la complejidad de la temática abordada en el trabajo, no siempre ha sido fácil mantener el enfoque del mismo, dado que fácilmente el seguimiento de una de las líneas de investigación hacía que se profundizase demasiado en algún ámbito que quedaba fuera de los objetivos del proyecto, por lo que ha sido necesario contar con la visión crítica de revisión del director del TFM, para volver al enfoque principal del trabajo. El poder conseguir la información sobre el estudio real sobre pacientes de cáncer de las características necesarias para trabajo, también ha representado un reto. Aun así, en base a los resultados

obtenidos, se puede concluir que los objetivos, tanto globales como específicos, que se marcaron en la definición del proyecto se han podido alcanzar en su totalidad. La metodología utilizada en cada uno de los ámbitos tratados en el trabajo y la rigurosidad en la dedicación de tiempo en cada una de las tareas definidas, ha ayudado de forma positiva a alcanzar los objetivos marcados en cada uno los hitos establecidos en la planificación del TFM.

Es evidente que se podrían definir infinidad de líneas futuras de trabajo como complemento o seguimiento de este trabajo, ya que como decía anteriormente, la complejidad de la temática tratada es muy alta. Un ámbito muy interesante a profundizar sería el de conseguir un realizar el mismo análisis de supervivencia de pacientes con cáncer de pulmón, pero con un mayor número de muestras válidas que den mayor evidencia estadística a las conclusiones extraídas. Es necesario emplear técnicas de NGS para conseguir la expresión diferencial de múltiples genes que ayuden a identificar y entender mejor el funcionamiento de las vías celulares en la progresión del cáncer y así la creación de terapias dirigidas más efectivas.

Capítulo 6

6. Glosario

AKT : Sinonimo de Protein kinase B (PKB), es una serine/threonine-specific protein kinase

ARAF : A-Raf proto-oncogene, serine/threonine kinase

A-Raf : A-Raf proto-oncogene, serine/threonine kinase

ASR : tasa estandarizada por edad ("age-standardised rate" en inglés)

AXL : Receptod de tyrosine kinase

BMK1 : del inglés "big MAP Kinase-1"

BRAF : B-Raf proto-oncogene, serine/threonine kinase

B-Raf : B-Raf proto-oncogene, serine/threonine kinase

CCR : Cáncer colorrectal

CDCP1 : del inglés "CUB domain containing protein 1"

COSMIC v84 : Cataloge of Somatic Mutations In Cancer, versión 84

CPCNP (NSCLC) : Carcinoma de pulmón célula no pequeña CPCNP (NSCLC - non-small-cell lung cáncer en inglés)

CRAF : del inglés "C-Raf proto-oncogene, serine/threonine kinase"

C-Raf/Raf-2 : Raf-1 proto-oncogene, serine/threonine kinase

CREB : "cAMP response element-binding", en inglés

ddCt : Algoritmo Delta-Delta-Ct (ddCt). Método de aproximación para determinar la expresión génica relativa con experimentos cuantitativos de PCR en tiempo real (qRT-PCR).

EGF : factor de crecimiento epidermico ("epidermal growth factor" en ingés)

EGFR : del inglés "Epidermal growth factor receptor"

ERK1 : sinónimo de mitogen-activated protein kinase 1

ERK1 : sinónimo de mitogen-activated protein kinase 3

ERK5 : sinónimo de mitogen-activated protein kinase 7

FTI : inhibidores de farnesiltrasferasa

GAP : del inglés "GTPase-accelerating proteins"

GDP : guanosine diphosphate

GEF : del inglés "Guanine Exchange Factor"

GGTI : inhibidores de geranil geraniltransferasa

Grb2 : del inglés "grown factor receptor-bound protein 2"

GTP : guanosine triphosphate

HES1 : del inglés "hairy and enhancer of split 1, (Drosophila)"

HR : riesgo relativo (" Hazard ratio", en inglés)

HRAS : del inglés "Harvey rat sarcoma viral oncogene homolog"

HVR : Domino hipervariable

JNK : cJun anino-terminal kinasa

KEGG : Kyoto Encyclopedia of Genes and Genomes

KRAS : del inglés "Kirsten rat sarcoma viral oncogene homolog"

LKB1 : sinónimo de serine/threonine kinase 11

MAP2K1 : del inglés "mitogen-activated protein kinase kinase 1"

MAP2K2 : del inglés "mitogen-activated protein kinase kinase 2"

MAPK : del inglés "Mitogen-activated protein Kinases"

MAPK/ERK : Vía de señalización también conocida como RAS-ERK (RAS-RAF-MEK-ERK)

MAPK1 : del inglés "mitogen-activated protein kinase 1"

MAPK3 : del inglés "mitogen-activated protein kinase 3"

NGF : factor de crecimiento nervioso ("nerve growth factor" en inglés)

NIH : National Cáncer Institute

NRAS : del inglés "neuroblastoma RAS viral (v-ras) oncogene homolog"

OMS : Organización Mundial de la Salud

PDEd : inhibidores de fosfodiesterasa d

PDGF : factor de crecimiento derivado de plaquetas

PDGFRA : del inglés "platelet derived growth factor receptor alpha"

RAS : superfamilia de proteínas de pequeñas GTPasas

RNA: ácido ribonucleico (“Ribonucleic acid” en inglés)

RTK : del inglés "Receptor tyrosine kinases"

Ser : aminoácido serina

SGL : Supervivencia Global

SLP : Supervivencia Libre de Progresión

SOS : del inglés "son of sevenles"

Src 2 – SH2 : del inglés Src homology 2

STAT3 : del inglés "signal transducer and activator of transcription 3"

Thr : aminoácido treonina

Tyr : aminoácido tirosina

WHO : World Health Organitation

YAP1 : del inglés "Yes associated protein 1"

Capítulo 7

7. Bibliografía

© Copyright 2016 RStudio Inc. (s.f.). R Markdown. Obtenido de https://rmarkdown.rstudio.com/

© The R Foundation. (s.f.). The R Project for Statistical Computing. Obtenido de http://www.cran.r-project.org/

Aronov, A. M. (2009). Structure-guided design of potent and selective pyrimidylpyrrole inhibitors of extracellular signal-regulated kinase (ERK) using conformational control. *J. Med. Chem, 52*, 6362-6368.

Baines, A., Xu, Dapeng, & Der, Channing J. (10 de 2011). Inhibition of Ras for cancer treatment: the search continues. *Future Medicinal Chemistry, 3*(14), 1787–1808. doi:10.4155/fmc.11.121

Bruce Alberts, A. J. (2002). *The Molecular Biology of the Cell 4th edition.* Garland Science Taylor & Francis Group.

Cargnello, Marie, & Roux, Philippe P. (Mar de 2011). Activation and Function of the MAPKs and Their Substrates, the MAPK-Activated Protein Kinases. *Microbiology and Molecular Biology Reviews : MMBR, 75*(1), 50-83. doi:10.1128/MMBR.00031-10

Contal, C., & O'Quigley, J. (1999). An application of changepoint methods in studying the effect of age on survival in breast cancer. *Computational Statistics and Data Analysis, 30*, 253-270.

COSMIC v84. (febrero de 2018). *Cataloge of Somatic Mutations In Cancer*. Recuperado el abril de 2018, de https://cancer.sanger.ac.uk/cosmic

Crespo Baraja, P. (2009). *Monografía XXIV. Redes de señalización y estrategias terapéuticas: La ruta RAS-ERK como diana tumoral.* Real Academia Nacional de Farmacia (España).

Dardis, C. (2016). Paquete surMisc de R - Miscellaneous Functions for Survival Data. Obtenido de https://cran.r-project.org/package=survMisc

David A. Eberhard et al. (2005). Mutations in the Epidermal Growth Factor Receptor and in KRAS Are Predictive and Prognostic Indicators in Patients With Non–Small-Cell Lung Cancer Treated With Chemotherapy Alone and in Combination With Erlotinib. *Journal of Clinical Oncology, 23*(25), 5900-5909. doi:10.1200/JCO.2005.02.857

Dirk Bokemeyer, A. S. (1996). Multiple intracellular MAP kinase signaling cascades. *Kidney International - EDITORIAL REVIEW, 49*, 187—1198.

Downward, J. (2003). Targeting RAS signalling pathways in cancer therapy. *Nature Reviews.Cancer, 3*(1), 11-22. doi:http://0-dx.doi.org.cataleg.uoc.edu/10.1038/nrc969

Flaherty, K. e. (2012). Combined BRAF and MEK inhibition in melanoma with BRAF V600 mutations. *New England Journal of Medicine, 367*(18), 1694-1703. doi:10.1056/NEJMoa1210093

FOAS: Foundation for Open Access Statistics - R Consortium. (s.f.). RStudio - Open source and enterprise-ready professional software for R. Obtenido de https://www.rstudio.com/

Gay, C. M., Balaji, K., & Byers, L. A. (2017). Giving AXL the axe: targeting AXL in human malignancy. *British Journal Of Cancer, 116*, 415-423. doi:10.1038/bjc.2016.428

Gil Moreno, M. (2018). *Carcinoma de pulmón No Celula Pequeña, con mutación de KRAS: Diferenciación y caracterización de subtipos, así como los diferentes mecanismos de resistencia, para elaboración de trataminetos dirigidos.* Tesis Doctoral.

Gilbert-Ross et al., M. (2017). Targeting adhesion signaling in KRAS, LKB1 mutant lung adenocarcinoma. *JCI insight, 2*(5), e90487. doi:10.1172/jci.insight.90487

Grabner et al., B. (2015). Disruption of STAT3 signalling promotes KRAS-induced lung tumorigenesis. *Nature Communications, 6*, 6285. doi:http://dx.doi.org/10.1038/ncomms7285

Hancock, J. F. (2003). Ras proteins: different signals from different locations. *Nature Reviews Molecular Cell Biology, 4*(Review), 373-384. doi:10.1038/nrm1105

Hatzivassiliou, G. e. (2012). ERK inhibition overcomes acquired resistance to MEK inhibitors. *Mol. Cancer Ther., 11*, 1143-1154.

HGNC. (2018). *HGNC - Hugo Gene Nomenclature Commitee*. Recuperado el abril de 2018, de National Human Genome Research Institute (NHGRI) and the Wellcome Trust (UK): https://www.genenames.org/

IARC - World Health Organitation. (2012). *International Agency for Research on Cancer.* Obtenido de Web Portal for International Cancer Research - GLOBCAN 2012: http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx

Kaelin, W. (2009). Synthetic lethality: a framework for the development of wiser cancer therapeutics. *Genome Medicine, 1*(10), 99. doi:https://doi.org/10.1186/gm99

Karnoub, A. E. (2008). Ras oncogenes: split personalities. *Nature reviews. Molecular cell biology, 9*(7), 517-531. doi:10.1038/nrm2438

KEGG. (Consultado en 2018). Kyoto Encyclopedia of Genes and Genomes- Pathway maps. Recuperado el 2018, de http://www.genome.jp/kegg/pathway.html

Kyung Kim, E., & Choi, E.-j. (2010). Pathological roles of MAPK signaling pathways in human diseases. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, 1802*(4), 396 - 405. doi:https://doi.org/10.1016/j.bbadis.2009.12.009

Lee et al., J. W. (2015). A Novel Small-Molecule Inhibitor Targeting CREB-CBP Complex Possesses Anti-Cancer Effects along with Cell Cycle Regulation, Autophagy Suppression and Endoplasmic Reticulum Stress. *PLoS ONE, 10*(4), e0122628. doi:10.1371/journal.pone.0122628

Lin et al., L. (2015). The Hippo effector YAP promotes resistance to RAF- and MEK-targeted cancer therapies. *Nature genetics, 47*(3), 250-256. doi:10.1038/ng.3218

Maosong Qi, & Elaine A. Elion. (2005). MAP kinase pathways. *Cell Science at a Glance*(118), 3569-3572. doi:10.1242/jcs.02470

Maraver et al., A. (2012). Therapeutic Effect of γ-Secretase Inhibition in KrasG12V-Driven Non-Small Cell Lung Carcinoma by Derepression of DUSP1 and Inhibition of ERK. *Cancer Cell, 22*(2), 222 - 234. doi:https://doi.org/10.1016/j.ccr.2012.06.014

Martinko et al., A. J. (2018). Targeting RAS-driven human cancer cells with antibodies to upregulated and essential cell-surface proteins. *eLife, 7*, e31098. doi:10.7554/eLife.31098

McCormick, F. (2015). KRAS as a Therapeutic Target. *Clinical cancer research : an official journal of the American Association for Cancer Research, 21*(8), 1797-1801. doi:10.1158/1078-0432.CCR-14-2662

Nan, X. T.-J. (2015). Ras-GTP dimers activate the Mitogen-Activated Protein Kinase (MAPK) pathway. *Proceedings of the National Academy of Sciences of the United States of America, 112*(26), 7996–8001. doi:http://doi.org/10.1073/pnas.1509123112

NIH - National Cáncer Institute. (2018). Obtenido de Cancer Statistics: https://www.cancer.gov/about-cancer/understanding/statistics

Nissan, M. H. (2013). ERK Pathway Inhibitors: How Low Should We Go? *CANCER DISCOVERY, 3*(7), 719–721.

Ohori, M. e. (2005). Identification of a selective ERK inhibitor and structural determination of the inhibitor-ERK2 complex.ERK inhibition overcomes acquired resistance to MEK inhibitors. *Biochem. Biophys. Res. Commun., 336*, 357-363.

Pao, William et al. (2005). KRAS Mutations and Primary Resistance of Lung Adenocarcinomas to Gefitinib or Erlotinib. *PLoS Medicine, 2*(1), 0057-0061. doi:10.1371/journal.pmed.0020017

Patricelli, M. e. (2016). Selective Inhibition of Oncogenic KRAS Output with Small Molecules Targeting the Inactive State. *Cancer Discovery, 6*(3), 316--329. doi:10.1158/2159-8290.CD-15-1105

Pearson, G. a.-e. (2001). Mitogen-Activated Protein (MAP) Kinase Pathways: Regulation and Physiological Functions. *Endocrine Reviews, 22*(2), 153-183. doi:10.1210/edrv.22.2.0428

Prior, I. A., & Lewis, P. D. (2012). A comprehensive survey of Ras mutations in cancer. *Cancer Research, 72*(10), 2457–2467. doi:10.1158/0008-5472.CAN-11-2612

Rubinfeld, H. &. (2005). The ERK cascade. *MOLECULAR BIOTECHNOLOGY, 31*(2), 151-174. doi:http://0-dx.doi.org.cataleg.uoc.edu/10.1385/MB:31:2:151

Samatar, A. A. (2014). Targeting RAS-ERK signalling in cancer: Promises and challenges. *Nature Reviews.Drug Discovery, 13*(12), 928-942. doi:http://0-dx.doi.org.cataleg.uoc.edu/10.1038/nrd4281

Seo, K.-Y., Jelinsky, S., & Loechler, E. (2000). Factors that influence the mutagenic patterns of DNA adducts from chemical carcinogens. *Mutation Research/Reviews in Mutation Research, 463*(3), 215-246. doi:http://www.sciencedirect.com/science/article/pii/S1383574200000478

Shen-Hsi Yang, A. D. (2013). MAP kinase signalling cascades and transcriptional regulation. *Gene, 513*(1), 1-13. doi:https://doi.org/10.1016/j.gene.2012.10.033

Siegel, R., Miller, K., & Jemal, A. (2017). Cancer statistics, 2017. *CA: A Cancer Journal for Clinicians, 67*(1), 7-30. doi:10.3322/caac.21387

Singer at al., C. A. (2011). Src mediates cytokine-stimulated gene expression in airway myocytes through ERK MAPK. *9*, 14. doi:10.1186/1478-811X-9-14

Steven A. Ahrendt et al. (2001). Cigarette smoking is strongly associated with mutation of the K-ras gene in patients with primary adenocarcinoma of the lung. *Cancer, 92*(6), 1525-1530. doi:10.1002/1097-0142(20010915)92:6<1525::AID-CNCR1478>3.0.CO;2-H

Takashima, A., & Faller, D. (2013). Targeting the RAS oncogene. *Expert Opinion on Therapeutic Targets, 17*(5), 507-531. doi:10.1517/14728222.2013.764990

Therneau, T. M. (2015). A Package for Survival Analysis in R - version 2.38. Obtenido de https://CRAN.R-project.org/package=surviva

Timothy P Garrington and Gary L Johnson. (1999). Organization and regulation of mitogen-activated protein kinase signaling pathways. *Current Opinion in Cell Biology, 11*(2), 211-218. doi:https://doi.org/10.1016/S0955-0674(99)80028-3

To, M. D. (2008). Kras regulatory elements and exon 4A determine mutation specificity in lung cancer. *Nature genetics, 40*(10), 1240-1244. doi: 10.1038/ng.211

UCSC. (2018). *Genome Browser* . Obtenido de University of California Santa Cruz: https://genome.ucsc.edu/

Victor. M. Valdespino-Gómez et al. (2015). Interacción de las vías de señalización intracelulares participantes en la proliferación celular: potencial blanco de

intervencionismo terapéutico. *Cirugía y Cirujanos, 83*(2), 165-174. doi:https://doi.org/10.1016/j.circir.2015.04.015

Wahlstrom, A. M. (2008). Inactivating Icmt ameliorates K-RAS–induced myeloproliferative disease. *Blood, 112*(4), 1357-1365. doi:https://doi.org/10.1182/blood-2007-06-094060

WHO - World Health Organitation. (2018). *World Health Organitation*. Recuperado el 2018, de Top 10 causes of death worldwide: http://www.who.int/mediacentre/factsheets/fs310/en/

Yoon et al., Y.-K. (2010). KRAS mutant lung cancer cells are differentially responsive to MEK inhibitor due to AKT or STAT3 activation: Implication for combinatorial approach. *Molecular Carcinogenesis, 49*, 353-362. doi:10.1002/mc.20607

Zimmermann, G. (2013). Small molecule inhibition of the KRAS-PDE interaction impairs oncogenic KRAS signalling. *Nature, 497*(638), 638-642. doi:http://dx.doi.org/10.1038/nature12205

Capítulo 8

8. Anexos

8.1 Anexo I. Informe de estudio poblacional del cáncer de pulmón

Se adjunta a la presente memoria del TFM el informe completo, en formato PDF, titulado *"Estudio poblacional de incidencia del cáncer de pulmón (Implicación de KRAS)"*, donde se detalla el estudio realizado sobre la incidencia del cáncer de pulmón respecto a otros tipos de cáncer en la población mundial y europea. La parte más esencial de este informe ha conformado el capítulo 3 de esta memoria.

El informe se ha realizado mediante RMarkdown (© Copyright 2016 RStudio Inc.) como informe dinámico programado en lenguaje R (© The R Foundation). En el mismo informe se adjunta como anexo el código R implementado para los diferentes estudios y gráficas incluidos en él.

8.2 Anexo II. Informe de análisis de supervivencia de cáncer NSCLC.

Adicionalmente, también se adjunta a la presente memoria del TFM el informe completo, en formato PDF, titulado "*Análisis Supervivencia de Cáncer NSCLC (Pacientes con mutación de KRAS)"*, donde se detallan el estudio de Supervivencia Libre de Progresión y Supervivencia Global, aplicando diferentes estrategias. En el informe se detalla el estudio original que ha servido de base para el presente trabajo y la posterior comparación de resultados. Se describen el pipeline del proceso de análisis realizado, los datos utilizados para el análisis, las diferentes funciones y estrategias aplicadas, así como los resultados numéricos y gráficos obtenidos en cada proceso. La parte más esencial de este informe ha conformado el capítulo 4 de esta memoria.

El informe se ha realizado mediante RMarkdown (© Copyright 2016 RStudio Inc.) como informe dinámico programado en lenguaje R (© The R Foundation). En el mismo informe se adjunta como anexo el código R implementado para los diferentes estudios y gráficas incluidos en él.