



Estudio poblacional de la mutación KRAS en pacientes con diagnóstico de cáncer de páncreas.

Carolina Arcones Martín.

Máster de Bioinformática y Bioestadística.

Biología del desarrollo, cáncer, biología molecular y farmacología.

Ivette Olivares Castiñeira.

Carles Ventura Royo.

2 de enero de 2019



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

B) GNU Free Documentation License (GNU FDL)

Copyright © 2019 Carolina Arcones.

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.

A copy of the license is included in the section entitled "GNU Free Documentation License".

C) Copyright

Título del trabajo:	Estudio poblacional de la mutación KRAS en pacientes con diagnóstico de cáncer de páncreas.
Nombre del autor:	<i>Carolina Arcones Martín</i>
Nombre del consultor/a:	<i>Ivette Olivares Castiñeira</i>
Nombre del PRA:	Carles Ventura Royo.
Fecha de entrega (mm/aaaa):	02/2019
Titulación:	<i>Máster de Bioinformática y Bioestadística.</i>
Área del Trabajo Final:	Biología del desarrollo, cáncer, biología molecular y farmacología.
Idioma del trabajo:	<i>Castellano.</i>
Palabras clave	<i>Cáncer de páncreas, mutación, KRAS.</i>
Resumen del Trabajo (máximo 250 palabras): <i>Con la finalidad, contexto de aplicación, metodología, resultados i conclusiones del trabajo.</i>	
<p>El cáncer de páncreas presenta una alta tasa de mortalidad en la población mundial. Se caracteriza por presentar en el 90% de los casos, la presencia de una mutación en la proteína KRAS.</p> <p>En el presente trabajo se realiza un análisis descriptivo y un estudio poblacional de las secuencias de pacientes con carcinoma de páncreas y otro tipo de carcinomas, para valorar las mutaciones que presentan en el exón 2, del codón 12, del cromosoma 12 del gen KRAS.</p> <p>Se realiza un estudio con 95 secuencias de ADN. 26 de estas secuencias pertenecen a carcinoma de páncreas y 69 a otro tipo de carcinomas. Todas las secuencias son descargadas de los bancos de datos genómicos.</p>	

Los resultados demuestran que las medias de las mutaciones en carcinoma de páncreas son superiores a los carcinomas de otra índole, obteniendo unas diferencias estadísticamente significativas en los pacientes con carcinoma de páncreas a través del tanto por cierto de las mutaciones según su longitud en la secuencia.

A través del estudio filogenético, se observan como se agrupan por enfermedades, apreciando la unión entre las muestras pancreáticas sin mutaciones y la dispersión en el árbol de las demás secuencias pancreáticas.

Principalmente, se ponen de manifiesto mutaciones de repetición en las muestras pancreáticas, merecedoras de futuros estudios, para poder observar su potencial influencia para desarrollar cáncer pancreático.

Se observan resultados de mutaciones de repetición interesantes para estudios posteriores. El estudio de estos cambios genéticos podría ayudar en el diagnóstico temprano, nuevas terapias más efectivas y mejora del pronóstico de los pacientes con carcinoma de páncreas.

Abstract (in English, 250 words or less):

Pancreas cancer has a high mortality rate among the world population. In 90 % of the cases appears as a consequence of the mutation in the KRAS protein.

In this work, a descriptive analysis a population study of the sequences of patients with pancreas carcinomas and other types of carcinomas have been carried out in order to evaluate the mutations that appear in the 2nd exon chromosome's 12 codon 12 the KRAS gene.

A study has been done with of DNA where 26 of these sequences belong to pancreatic carcinoma and 69 belong to other types of carcimomas. All the sequences are discharged from the bank of genomics information.

The results show that the averages of the mutations in pancreas carninomas are higher than in other kinds of carcinomas obtaining statistically significant differences

The results show that the averages of the mutations in pancreas carcinomas are higher than in other kinds of carcinomas obtaining statistically significant differences in patients with pancreatic carcinoma through a percentage of the mutations depending on the length of the sequence.

In the phylogenetic study, the mutations are grouped by diseases, appreciating the union between the pancreatic samples without mutations and the dispersion in the tree of the other pancreatic sequences, where there is more diversity in genetic changes.

Repetitive mutations are shown mainly in pancreatic samples. These results demonstrate their potential influence to develop pancreatic cancer worthy of further studies.

The study of these genetic changes could help in early diagnosis, new more effective therapies and improve the prognosis of patients.

ÍNDICE

1. INTRODUCCIÓN.....	8
1.1. Contexto y justificación del Trabajo.....	8
1.2. Objetivos del Trabajo.....	11
1.3. Enfoque y método seguido.....	12
1.4. Planificación del Trabajo.....	14
1.5. Breve resumen de los productos obtenidos.....	20
1.6. Breve descripción de los otros capítulos de la memoria.....	21
2. DESARROLLO.....	22
2.1. Proteína KRAS.....	22
2.2. Población en estudio.....	24
2.3. Mutaciones en la proteína KRAS.....	28
2.4. Análisis del gen KRAS.....	28
2.5. Fragmentación de las secuencias de ADN.....	29
2.6. Análisis de los cambios genéticos.....	35
2.7. Análisis descriptivo.....	37
2.8. Estudio poblacional.....	44
3. RESULTADOS.....	56
3.1. Resultados del análisis de las mutaciones en las secuencias.....	56
3.2. Resultados del análisis descriptivo.....	57
3.3. Resultados del análisis filogenético.....	58
4. CONCLUSIÓN.....	60
4.1. Futuros estudios.....	61
5. GLOSARIO.....	62
6. BIBLIOGRAFÍA.....	63
7. ANEXOS.....	67

LISTADO DE FIGURAS

Figura 1. TAC abdominal donde se aprecia un carcinoma de páncreas.....	9
Figura 2. Esquema vía EGFr, representando el gen KRAS.....	22
Figura 3. Localización del gen KRAS.....	23
Figura 4. Fragmentos del proyecto.....	26
Figura 5. Servidor genómico UCSC.....	27
Figura 6. Mutación G12D.....	28
Figura 7. Fragmento 1 visionado en BioEdit.....	30
Figura 8. Fragmento 2 visionado en BioEdit.....	31
Figura 9. Fragmento 3 visionado en BioEdit.....	31
Figura 10. Fragmento 4 visionado en BioEdit.....	32
Figura 11. Fragmento 5 visualizado en BioEdit.....	34
Figura 12. Gráfica de los datos recabados.....	39
Figura 13. <i>summary()</i> de los datos en R.....	40
Figura 14. Resultados obtenidos con la función <i>describeBy()</i> en R.....	41
Figura 15. Resultados obtenidos con la función <i>descrbeBy()</i> en R.....	41
Figura 16. Boxplot del número de mutaciones.....	42
Figura 17. Boxplot con el tanto por cierto de mutaciones.....	43
Figura 18. Árbol filogenético Neighbour-Joining del fragmento 1.....	46
Figura 19. Árbol filogenético Maximum Likelihood del fragmento 1.....	47
Figura 20. Árbol filogenético Neighbour-joining del fragmento 2.....	48
Figura 21. Árbol filogenético Maximun Likelihood fragmento 2.....	49
Figura 22. Árbol filogenético Neighbour-Joining del fragmento 3.....	50
Figura 23. Árbol filogenético Maximum Likelihood del fragmento	51
Figura 24. Árbol filogenético Neighbour-Joining del fragmento 4.....	52
Figura 25. Árbol filogenético Maximum Likelihood del fragmento 4.....	53
Figura 26. Árbol filogenético Neighbour-Joining del fragmento 5.....	54
Figura 27. Árbol filogenético Maximum Likelihood del fragmento 5.....	55

LISTADO DE TABLAS

TABLA 1. Tabla de planificación del proyecto.....	15
TABLA 2. Tabla del diagrama de Gantt.....	18
TABLA 3. Tabla resumen de las secuencias.....	25
TABLA 4. Tabla de secuencias pancreáticas y no pancreáticas.....	26
TABLA5. Tabla número de secuencias y mutaciones.....	35
TABLA 6. Tabla de información de ADN y su longitud.....	36
TABLA 7. Representación de las muestras pancreáticas y no pancreáticas.....	38

1. INTRODUCCIÓN.

1.1. Contexto y justificación del Trabajo.

El carcinoma de páncreas está causando hasta 330.000 muertes al año en el mundo (E. Pelosi et al. 2017). Tiene una alta mortalidad, llegando a alcanzar el 90%, en la mayoría de carcinomas ductales. Incluso en los estadios tempranos, cercanos al 20% de los diagnósticos, con tumores localizables y resecables, es decir, que se puede extraer por medio de cirugía, el pronóstico tampoco es bueno. Los pacientes tienen una supervivencia de alrededor de 18 meses, menor del 5% a los cinco años (S. Sho et al. 2017). La dificultad de su diagnóstico está causada por la ausencia de sintomatología en estadios tempranos, unido a la agresividad biológica característica de este cáncer que imposibilita un diagnóstico precoz.

El problema de su alta mortalidad también radica en su gran capacidad de difusión local y en la capacidad de producir metástasis a distancia. Esta disposición, unida su tardío diagnóstico en el 80% de los casos, hace que posea un evidente mal pronóstico (E. Pelosi et al. 2017, M. Hidalgo et al. 2010).

En la Figura 1, podemos observar un TAC abdominal con contraste, donde vemos la localización del carcinoma de páncreas y su cercanía con otros órganos, situación que ayuda a su extensión local.



Figura 1. TAC abdominal donde se aprecia un carcinoma de páncreas. Ref. A. Utrillas et al. 2018.

Las causas que favorecen el carcinoma de páncreas son desconocidas. Algunos autores hablan del consumo de tabaco, alcohol, café o incluso el ácido acetilsalicílico. Del mismo modo, tienden más a padecer cáncer de páncreas pacientes con antecedentes de diabetes, pancreatitis crónica, dietas con un contenido alto en grasas y colecistectomía. Asimismo, hay que tener en cuenta, que el 5-10% de los pacientes tienen antecedentes familiares, cuando existen más de cuatro miembros en la familia diagnosticados con cáncer de páncreas, se alcanza un riesgo 57 veces más elevado de padecer este tipo de cáncer (M. Hidalgo et al. 2010).

En resumen, el adenocarcinoma de páncreas tiene un diagnóstico tardío por la ausencia de síntomas en estadios tempranos, además de poseer una alta agresividad biológica y baja supervivencia en estadios tempranos de tan solo 18 meses. Estos datos no se han

modificado desde hace 30 años (E. Pelosi et al. 2017). Como consecuencia de estas particularidades, se hacen necesarios estudios que aumenten el conocimiento sobre la enfermedad, y en particular, sobre la mutación genética KRAS que se encuentra sobreexpresada en el 90% de los diagnósticos (E. Pelosi et al. 2017).

Según la bibliografía recabada en el trabajo de fin de máster, la mutación genética KRAS se ha detectado en tumores de páncreas, colon y pulmón, siendo en el páncreas donde peor pronóstico tienen (E. Pelosi et al. 2017; A. M. Krasinskas et al. 2013; P. A. López Penza et al. 2017).

Los datos expuestos sugieren la necesidad de realizar estudios de la mutación genética en la proteína KRAS. Estudios enfocados a conseguir terapias más personalizadas y dirigidas mediante fármacos más efectivos que ayuden a mejorar el tratamiento. De esta forma, poder aumentar la supervivencia de los pacientes, además de conseguir un diagnóstico precoz.

En los últimos años, ha habido un aumento de los datos biológicos, sobre todo de los datos genéticos y genómicos. Este aumento de datos ha derivado en un gran avance de las técnicas bioinformáticas, imprescindibles para poder solventar la problemática del análisis del gran número de datos que se obtienen en los experimentos en el laboratorio.

El estudio de los cambios en las secuencias del gen KRAS se realiza a través de herramientas bioinformáticas, con programas tipo BioEdit o Mega, o incluso herramientas online. Herramientas que han conseguido un alto grado de fiabilidad, aumentando la exactitud de los resultados y minimizando el tiempo de análisis.

La bioinformática, no solo ha facilitado el análisis de datos, igualmente ha posibilitado el almacenaje, mantenimiento, interpretación y relación de los mismos, en bancos de datos y servidores genéticos. En este proyecto se obtienen secuencias de otros autores almacenadas en los bancos de datos genómicos.

A continuación, y una vez obtenidas las secuencias, se hace un análisis e interpretación de los resultados obtenidos en el estudio. Se ha intentado recabar la mayor información sobre las secuencias obtenidas en pacientes con carcinoma de páncreas y pacientes no pancreáticos.

1.2 Objetivos del Trabajo.

Objetivos generales:

1. Análisis descriptivo de las secuencias encontradas en las bases de datos de genes de la proteína KRAS en la región de interés.
2. Estudio poblacional de las secuencias descargadas de los bancos de datos.

Objetivos específicos:

1. Primer objetivo:
 - 1.1. Extraer las anotaciones de RefSeq de la región codificante de la mutación KRAS.
 - 1.2. Buscar de secuencias con la mutación KRAS de pacientes diagnosticados con carcinoma de páncreas, publicadas en los bancos de datos de genes.
 - 1.3. Búsqueda de secuencias de otro tipo de carcinomas en la misma región.
 - 1.4. Estudiar las secuencias en la región de interés.
 - 1.5. Análisis descriptivo de los resultados obtenidos.
2. Segundo objetivo:
 - 2.1. Sobre el estudio anterior y con las secuencias ya anotadas, alineadas y preparadas, se hará un estudio poblacional con las secuencias descargadas de los bancos de datos, a través de filogenética.

1.3 Enfoque y método seguido.

Este estudio tiene como objeto el estudio de las mutaciones en el gen KRAS en pacientes con carcinoma de páncreas y pacientes con otro tipo de carcinomas. Para ello, se hace un análisis de la región de interés en secuencias encontradas en los bancos de datos genómicos. Se han valorado los cambios en las secuencias comparándolas entre pacientes con carcinomas pancreáticos y pacientes con otro tipo de carcinomas.

Es necesario adquirir un mayor conocimiento sobre el gen KRAS, por la implicación que tiene en algunos cánceres como en el de páncreas. Así pues, estudios que amplíen el conocimiento del gen KRAS, se hacen necesarios para intentar resolver su alta letalidad, debido a su agresividad biológica y gran capacidad de producir metástasis. Nuevas informaciones, ayudarían a desarrollar mejoras en los tratamientos con fármacos más personalizados y dirigidos que aumenten la calidad de vida y la supervivencia del paciente, mejorando su pronóstico.

El estudio que se realiza en el trabajo de fin de máster consta de un estudio descriptivo y otro poblacional de las secuencias de ADN descargadas de las bases de datos genómicas.

Este proyecto, se enfoca desde la aplicación de herramientas bioinformáticas, con Biostrings, Bioconductor y DESeq2 de R, y programas como BioEdit o Mega. También, el uso de servidores genómicos, bancos de datos y otros programas en línea como NCBI, MUSCLE, UCSC Genome, GenBank, BLAST, etc. Para aplicarlos al estudio de los cambios genéticos en el gen KRAS de secuencias de carcinomas pancreáticos y no pancreáticos.

Otro punto a tener en cuenta de igual manera, es la dificultad que supone el diagnóstico precoz por la ausencia de síntomas en el paciente. Un mayor conocimiento de la enfermedad y de los factores que la favorecen, ayudarían a un diagnóstico más rápido aumentando la supervivencia del paciente.

La metodología a seguir:

- Después de definir el proyecto, se harán las anotaciones del RefSeq, es decir, se buscará la secuencia de ADN consenso de la proteína KRAS en formato FASTA, o secuencia de referencia. La búsqueda de la RefSeq, que se utiliza como guía, se realiza en los bancos de datos de genes como, por ejemplo: UCSC, UniGene o GenBank.
- Descargar de todas las secuencias de ADN en formato FASTA de las bases de datos genómicas que corresponden con la zona del gen KRAS del estudio, de pacientes con carcinomas pancreáticos y no pancreáticos.
- Pipelines de las anotaciones de la región genómica.
- Evaluación de las secuencias de ADN en formato FASTA haciendo una primera visualización en el programa informático BioEdit o Mega como control de calidad.
- Alineamiento de las secuencias descargadas en las bases de datos, para analizarlas y compararlas con EMBL de MUSCLE, ClusterW y BioEdit.
- Estudio de las secuencias visualizando las mutaciones o cambios que se producen en la región. Estudio de las relaciones entre secuencias, las similitudes o diferencias encontradas en el gen KRAS.
- Análisis descriptivo de las secuencias del gen KRAS obtenidas en los bancos de datos a través de herramientas informáticas BioEdit, MEGA 5 y 7.
- Análisis descriptivo de los datos de las secuencias de ADN que se realiza a través de R.
- Estudio poblacional a través de la clasificación de la información molecular que permite establecer relaciones entre las secuencias usando MEGA 5 y 7, MUSCLE, Tree View, entre otras herramientas bioinformáticas. Construcción de árboles filogenéticos por los métodos de Neighbor-joining y Máxima verosimilitud, donde se hace una representación gráfica de las agrupaciones entre las secuencias de ADN basados en distancias.

1.4 Planificación del Trabajo.

El proyecto se desarrolla según la planificación de la tabla 1. El trabajo de fin de máster se ha elaborado según los plazos previsto con entregas de resultados parciales divididas en PEC. Durante todo el proyecto se han tenido en cuenta artículos publicados relacionados con el carcinoma de páncreas y las mutaciones en el gen KRAS. Actualizaciones que pueden mejorar el desarrollo del proyecto.

PEC	Actividad	Fecha inicio	Fecha fin	Tareas
PEC0	Definición de los contenidos	19/09/2018	01/10/2018	Lectura de artículos.
				Elección del TFM.
				Problemática a resolver.
				Definir objetivos generales.
PEC1	Desarrollo del TFM	02/10/2018	15/10/2018	Lectura de artículos.
				Concretar objetivos más específicos.
				Desarrollo del plan de trabajo.
				Investigar las herramientas informáticas.
				Instalación de las herramientas informáticas.
PEC2	Desarrollo del TFM	16/10/2018	19/11/2018	Realización de la memoria correspondiente.
				Lectura de artículos.
				Búsqueda de secuencias.
				Editar secuencias.
				Alinear secuencias.
				Análisis de datos con las herramientas informáticas.
PEC3	Desarrollo del TFM	20/11/2018	17/12/2018	Informe del seguimiento.
				Lectura de artículos.
				Relación entre secuencias.
				Análisis descriptivo.
				Estudio poblacional.
				Interpretación de los resultados.
PEC4	Redacción de la memoria	18/12/2018	02/01/2018	Informe de seguimiento.
				Redactar la memoria definitiva del TFM.
PEC5a	Realización de la presentación	03/01/2019	10/01/2019	Entrega de la memoria definitiva.
				Realización de la presentación.
PEC5b	Defensa del TFM	14/01/2019	23/01/2019	Exposición.

TABLA 1. Tabla de planificación del proyecto, donde se incluyen las actividades dentro de las fechas establecidas de cada PEC

PEC0: 19/09/2018 – 01/10/2018

- Lectura de artículos para la toma de decisiones sobre el trabajo de fin de máster.
- Sobre la literatura leída, selección de un tema de interés para realizar el Trabajo de fin de máster.
- Selección de la problemática a resolver.
- Definir objetivos.

- Realizar la memoria justificativa.

PEC1: 02/10/2018 – 15/10/2018

- Lectura de artículos.
- Concretar los objetivos definidos en la PEC0.
- Investigar las herramientas que van a ser necesarias para el análisis de las secuencias que se obtienen de los bancos de datos.
- Instalación de las herramientas informáticas.
- Elaborar la memoria correspondiente.

PEC2: 16/10/2018 – 19/11/2018

- Instalar el Paquete R, BioEdit, TreeView, y demás herramientas informáticas comprobando que su funcionamiento sea asequible y eficaz.
- Seguir leyendo artículos relacionados
- Búsqueda de secuencias con la mutación KRAS en artículos publicados en Pubmed.
- Búsqueda de secuencias de ADN en formato FASTA, desde la colección de secuencias anotadas en el NCBI, usando BLAST y obteniendo secuencias homólogas de pacientes con carcinoma de páncreas, mediante el servidor genómico UniGene, UCSC, Biomart o Genbank. Estas herramientas bioinformáticas nos permiten alinear y buscar secuencias de la misma región genómica.
- Editar las secuencias en formato FASTA con el editor BioEdit, que permite abrir cromatogramas para poder editar y corregir las secuencias, haciendo un análisis visual como control de calidad.
- Pipelines de Anotación de la localización genómica
- Extraer las anotaciones de RefSeq de la región codificante del gen KRAS.
- Alinear y comparar las secuencias de ADN encontradas, con programas como ClustalW Omega. Análisis de los alineamientos de las secuencias de ADN, haciendo un examen visual y observación sus diferencias y similitudes. Anotación de todos los datos encontrados y resultados que nos proporcionen estas herramientas informáticas.
- Desarrollar un informe de memoria con el trabajo realizado hasta el momento, los problemas o retrasos que se hayan podido dar y los resultados obtenidos.

PEC3: 20/11/2018 – 17/12/2018

- La herramienta Table Browser del navegador del UCSC apoya a ejecutar múltiples acciones que involucren a las secuencias, como calcular la correlación entre las secuencias de ADN y las anotaciones servidas por el proyecto Ref Seq, calculando la similitud entre las secuencias del experimento.
- Estudio descriptivo de los resultados obtenidos con Bioconductor, RNA-seq, etc. En lenguaje R.
- Construcción de árboles filogenéticos para realizar el estudio poblacional, que nos muestre una visualización interactiva y gráfica de los datos de las secuencias de ADN genómicos basados en distancias. Uso de los métodos de Neighbor-joining y Máxima verosimilitud basados en árboles jerárquicos. Es una representación gráfica de los resultados de la agrupación de las secuencias.
- Interpretación de los resultados obtenidos.
- Desarrollar un informe de memoria con el trabajo realizado hasta el momento, los problemas o retrasos que se hayan podido dar.
- Continuar ampliando bibliografía para actualizar la información que precise.

PEC4: 18/12/2018 – 02/01/2019

- Elaboración de la memoria del Trabajo de Fin de Máster con los resultados obtenidos durante las PEC.

PEC5a: 03/01/2018 – 10/01/2019

- Realizar las diapositivas para la presentación del Proyecto de Fin de Máster.

PEC5b: 14/01/2019 – 23/01/2019

- Defensa del Trabajo de Fin de Máster.

Para concluir, en todas las PEC se ha realizado una memoria justificativa de las tareas realizadas, incluyendo las novedades que se han introducido en el proyecto y los cambios necesarios para poder solventar problemáticas. En la tabla 1, podemos ver el calendario de actividades realizadas con más detalle de las actividades desarrolladas y los tiempos establecidos.

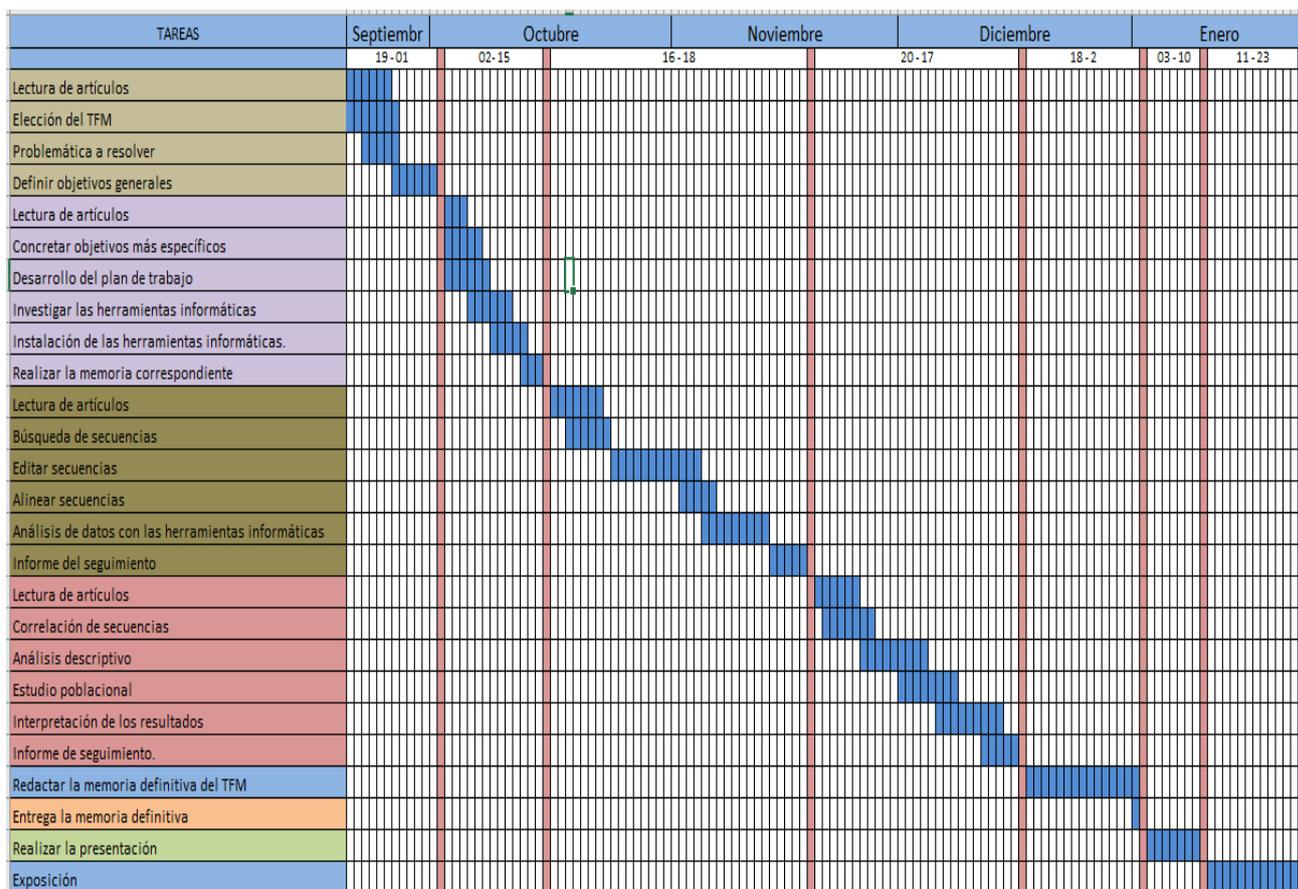


TABLA 2. Tabla del diagrama de Gantt donde se esquematiza la planificación del proyecto, se incluyen las actividades relacionando el tiempo de dedicación de cada tarea.

Las tareas que se han realizado:

19/09/2018 – 01/10/2018:

- Elección del trabajo de fin de máster y su problemática a resolver.
- Elección de los objetivos generales.
- Informe de seguimiento de la PECO

02/10/2018 – 15/10/2018:

- Concretar objetivos más específicos.
- Desarrollar el plan de trabajo.
- Investigación de las herramientas informáticas.
- Informe de seguimiento de la PEC1.

16/10/2018 – 19/11/2018:

- Búsqueda de secuencias de ADN.
- Editar y alinear secuencias de ADN.
- Análisis de datos de las secuencias descargadas.
- Informe de seguimiento de la PEC2.

20/11/2018 – 17/12/2018.

- Correlación entre secuencias.
- Análisis descriptivo.
- Estudio poblacional.
- Interpretación de los resultados.
- Informe de seguimiento de la PEC3.

18/12/2018 – 02/01/2019.

- Redacción de la memoria definitiva del trabajo de fin de máster.
- Entrega de la memoria definitiva.

03/01/2019 – 10/01/2019.

- Realización de la presentación de las diapositivas.

14/01/2019 – 23/01/2019

- Exposición del trabajo de fin de máster.

En la tabla 2, se ha desarrollado un diagrama de Gantt para representar las tareas y los tiempos de dedicación que se establecen en cada PEC.

1.5 Breve resumen de los productos obtenidos.

A continuación, se hace un pequeño resumen de los resultados obtenidos en el trabajo final del Máster de Bioinformática y Bioestadística:

- Observación los diferentes cambios genéticos de las secuencias genómicas de pacientes con carcinoma de páncreas y otros tipos de carcinomas.
- Analizar comparativamente de las secuencias alineadas en los diferentes programas bioinformáticos.
- Elaboración de un estudio descriptivo de las muestras con R como herramienta informática.
- Construcción de árboles filogenéticos a través de dos métodos diferentes.
- Resultados y conclusiones de los productos obtenidos en los diferentes puntos.

1.6 Breve descripción de los otros capítulos de la memoria.

Dentro del punto de desarrollo se ha elaborado diferentes puntos según las necesidades del proyecto.

1. Proteína KRAS

Se describe la proteína KRAS en la vía EGFR, su localización y funciones dentro de la célula. Además de la potencial influencia del oncogén KRAS en el carcinoma de páncreas.

2. Población en estudio

En este punto se describe las dos poblaciones en estudio, los pacientes que padecían carcinoma de páncreas y los pacientes que padecían otro tipo de carcinoma. Se extraen secuencias de los servidores genómicos de ambas poblaciones para su posterior análisis.

3. Mutaciones en la proteína KRAS

Descripción de las mutaciones que pueden dar lugar a padecer carcinoma de páncreas ubicando la región a estudio dentro del cromosoma 12 del gen KRAS.

4. Análisis del gen KRAS

Visualización del tipo de secuencias que se incluyen en el proyecto. Alineamiento de las secuencias encontradas en los bancos de datos genómicos con ayuda de diferentes programas bioinformáticos.

5. Fragmentación de las secuencias de ADN.

Fragmentación del alineamiento en 5 fragmentos para poder optimizar todas las secuencias descargadas de los bancos de datos. Cuanto mayor es el fragmento, menos muestras participan en la alineación. En esta parte, se ponen de manifiesto los cambios genéticos que tienen las secuencias de ADN.

6. Análisis de los cambios genéticos.

Visualización y argumentación de los diferentes cambios genéticos en las secuencias de ADN que se aprecian en las secuencias. Observación entre las diferencias y similitudes que presentan las secuencias a nivel genético.

7. Análisis descriptivo

Separación de las muestras en dos grupos, los pacientes con carcinoma de páncreas y los pacientes con otro tipo de carcinomas. A partir de aquí, se hace un estudio a través de R, para comprobar si existen diferencias estadísticamente significativas entre las dos poblaciones de pacientes con cáncer.

8. Estudio poblacional

Estudio poblacional de las secuencias de ADN a través de filogenética. Observación en los árboles de las distancias que existen entre ellas y relación entre los carcinomas que poseen y las agrupaciones en las ramas.

9. Resultados y Conclusión

Argumentación de los resultados y conclusiones de los productos obtenidos en los análisis que se realizan a lo largo del trabajo de fin de máster.

2. DESARROLLO.

2.2 Proteína KRAS.

La proteína KRAS que se encuentra en la vía del gen EGFR, se sitúa en la membrana plasmática de la célula y tiene como función regular la actividad de GTP-hidrolasa. Se puede encontrar activada unida a GTP o inactiva unida a GDP. Representado en la figura 2.

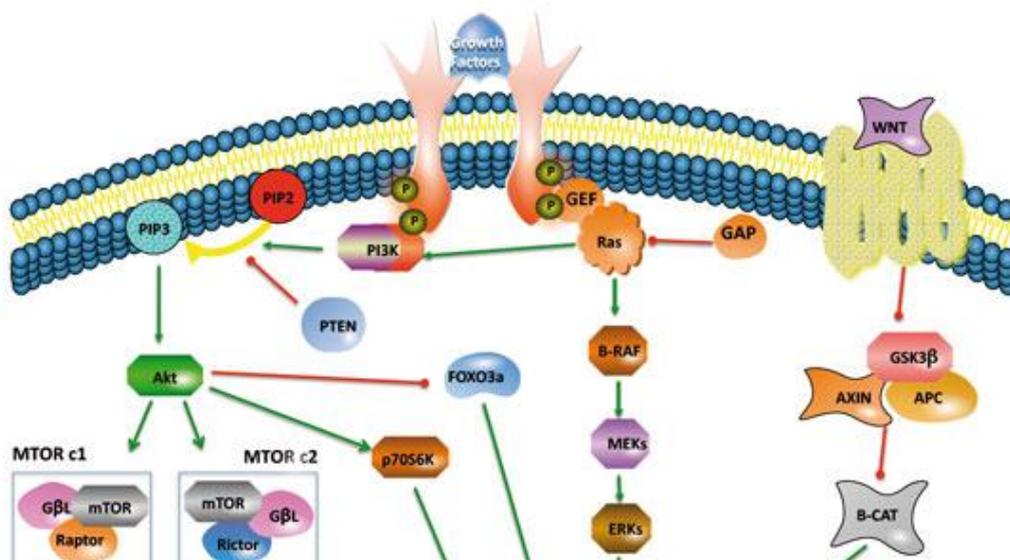


Figura 2. Esquema de la membrana plasmática de la célula, vía EGFR, representando el gen KRAS. Ref: G Argilés et al. 2012.

La proteína KRAS tiene un bajo peso molecular con alrededor de 45.000 pares de bases, que se localizan en el brazo corto del cromosoma 12 (12p.12.1). La mutación del gen KRAS, G12D/G12V se localiza en el exón 2, codón 12 del cromosoma 12 (S. Sho et al. 2017). La localización del gen KRAS está representada en la figura 3.

KRAS

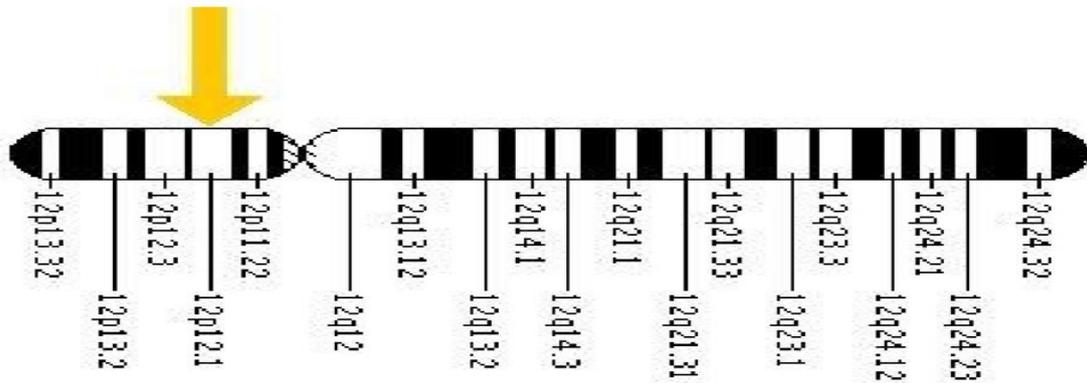


Figura 3. La flecha señala la localización del gen KRAS dentro del brazo corto del cromosoma 12. <http://cancerdecolon.blogspot.com>.

Una mutación somática sucede como consecuencia de un cambio en las secuencias de ADN de nucleótidos, ya sea un cambio de aminoácido, una delección o inserción. Si se da este proceso, y fallan los mecanismos de corrección de la célula, las células contendrán esa mutación que se heredará, creándose dos líneas celulares con genotipos diferentes. La primera línea celular serán las células sanas sin mutación y la segunda serán las células mutadas que formarán el tumor.

La presencia de células que contienen el gen mutado KRAS, juega un papel importante en el mal pronóstico de la enfermedad, por el aumento de la supervivencia y división de las células mutadas.

En este trabajo de fin de máster, se hace un estudio de las mutaciones somáticas que se producen en el gen KRAS, precursoras del carcinoma de páncreas. Pero, como la mutación somática en la proteína KRAS que está ligada a varios tipos de cáncer, se hace un estudio de dos poblaciones. La primera población es la que tiene afectación pancreática y la segunda es la que tiene afectación en otro tipo de carcinomas no pancreáticos, comparando las secuencias de ADN y viendo otras mutaciones que se puedan producir.

Cuando se produce una mutación en la proteína KRAS, se altera la actividad de la GTP-hidrolasa enviando señales anómalas mitogénicas a la célula rompiendo el equilibrio de

la proliferación celular. Es decir, la alteración de la hidrólisis de GTP produce la activación del oncogén KRAS, transformando la célula en malignas y aumentando la supervivencia y la proliferación de las células anómalas formadas (S. Yachida et al. 2010).

Existen varios factores que favorecen la agresividad del adenocarcinoma de páncreas, como la gran capacidad de reordenamiento, por el anómalo control celular y la disfunción de los telómeros. Estos procesos ocurren en la fase G1-S del ciclo celular (P. J. Campbell et al. 2010).

Asimismo, la activación del oncogén KRAS induce la inactivación de TP53, CDK2A y SMAD4, que son supresores tumorales. Esta situación favorece el descontrolado crecimiento del carcinoma. El oncogén KRAS, se puede encontrar en el tumor primario y en las metástasis a distancia en otros órganos secundarios (S. Yachida et al. 2010).

Se trata de un cáncer con una gran heterogeneidad genética con mutaciones como KRAS, APC, TP53, SMAD4 y CDKN2A que favorecen su avance (M. Hidalgo et al. 2010). Este estudio se centra en las mutaciones de la proteína KRAS porque se ha demostrado que dicha mutación está relacionada con el diagnóstico tardío, y un peor pronóstico de la enfermedad (A. M. Krasinkas et al. 2013). Las mutaciones en KRAS se pueden encontrar en lesiones iniciales del tumor primario y en las metástasis en otros órganos a distancia (P. A. López-Penca et al. 2017), encontrándose en el 90% de los diagnósticos de carcinoma (E. Pelosi et al. 2017).

2.2 Población en estudio.

Para empezar, mencionaremos el tipo de pacientes que han participado en el proyecto. Se inicia por la búsqueda de secuencias en los bancos de datos genómicos. Para comenzar, se ha buscado una secuencia RefSeq en el UCSC Genome, se trata de NM004985.4. Las secuencias RefSeq son un consenso de secuencias homólogas que se encuentran en los bancos de datos genómicos y se utilizan como guía.

Las secuencias seleccionadas pertenecen al exón 1 y 2 del codón 12 del cromosoma 12 del gen KRAS. Se pretenden obtener secuencias que contengan la mutación G12D Aspartato y G12V Valina (I. Roa et al 2013, S. Sho et al. 2017). Según los estudios realizados hasta el momento, otorga una mayor agresividad biológica al carcinoma de páncreas.

Las secuencias pertenecientes a carcinoma de páncreas descargadas de los servidores genéticos como NCBI, UCSC Genome, Gene Card, etc. no contienen la zona de la mutación, situación que nos da la oportunidad de descargar secuencias de otros tipos de carcinomas, como observamos en la tabla del anexo 1.

A partir de este momento, el proyecto consta dos poblaciones diferentes, los pacientes con carcinomas pancreáticos y los pacientes con carcinomas no pancreáticos. En la siguiente tabla se hace un resumen de las secuencias encontradas.

Type_cancer	Sample	pb/media	mut	%mut
RefSeq	1	442	0	0
Bladder	24	101	0	0
Brain_mut	1	442	1	0,68
Brain	1	108	0	0
Cervical	3	95,33	0	0
Colon_mut	1	108	1	0,93
Colon	5	55,2	0	0
Gallbladder	1	108	1	0,93
Lung	4	197	0	0
Lung_mut	5	255,6	13	0,93
lekemia_mu	1	108	1	0,93
Osteosarcoma	1	108	2	1,85
Pancreatic	7	66,85	0	0
Pancreatic_mut	19	60	107	9,39
Prostate	1	442	0	0
Sarcoma	1	108	0	0
Testis	1	442	0	0
Unknown	11	257,9	0	0
Unknown_mut	7	419,9	63	2,05

TABLA 3. Resumen de las secuencias encontradas en las bases de datos genómicas. En ella, se especifica las secuencias recogidas de cada tipo de cáncer, la longitud media, y el número de mutaciones.

En los no pancreáticos encontramos secuencias de osteosarcoma, sarcoma, vesícula, leucemia mieloide aguda, carcinoma de vejiga, de cervix, testicular, cerebro, pulmón, colorrectal, protático y secuencias con carcinomas desconocidos, el GenBank no ofrece más datos como vemos en la tabla 3.

	Seq	Mut
Pancreatic	7	0
Pancreatic_mut	19	107
NoPancreatic	51	0
NoPancreatic_mut	18	84

TABLA 4. Secuencias pancreáticas y no pancreáticas y sus mutaciones correspondientes.

Se descargan un total de 95 secuencias, dentro de ellas, 26 son secuencias de pacientes con carcinoma pancreático y 69 de no pancreáticos, donde se realiza el estudio de mutaciones resumido en la tabla 4.

Para un mayor aprovechamiento de las secuencias se divide en 5 fragmentos representados en la figura 4. Cuanto mayor es el fragmento, menor es el número de secuencias que pueden participar porque no tienen, en su mayoría, un tamaño adecuado y no cubren la zona adecuada en la alineación. Para localizar cada fragmento se toma como referencia la secuencia RefSeq.



Figura 4. Fragmentos en los que se dividen las secuencias encontradas en los bancos de datos genómicos.

Frag 1: Se localiza entre el 193-292, con 100pb, tomando como referencia para la localización la secuencia RefSeq. En este fragmento se hace el estudio de las mutaciones en G12D y G12V correspondiente a la localización 228 según la secuencia RefSeq. Este fragmento contiene 32 secuencias.

Frag 2: Se localiza entre 342-443, con 101pb, con referencia a la secuencia RefSeq. En él se consiguen alinear 48 secuencias, siendo uno de los fragmentos con más secuencias incluidas.

Frag 3: Se localiza entre 303-480, con una longitud de 178pb. Dentro de este fragmento engloba el Frag2, pero al ampliar el fragmento, disminuyen en número de secuencias que participan, siendo 24.

Frag 4: Se localiza entre el 193-634, con una extensión de 442pb. En este fragmento se encuentran incluidos todos los demás fragmentos. Se intenta estudiar un fragmento con la mayor longitud posible, pero tiene como consecuencia, la inclusión de solo 19 secuencias.

Frag 5: Se localiza entre las posiciones 243-301, con una longitud de tan solo 60 pares de bases. Este fragmento es importante porque en él se incluyen las secuencias de pacientes con carcinoma de páncreas. Comprende 49 secuencias y entre ellas las 26 de secuencias pancreáticas.

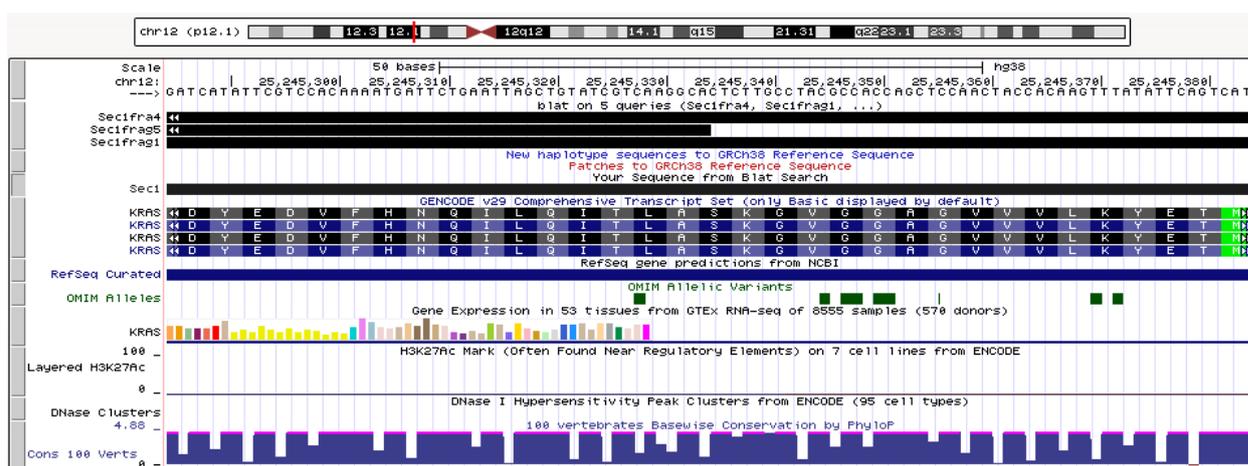


Figura 5. Localización del fragmento 1, 4 y 5 dentro del codón 12 del cromosoma 12 del gen KRAS en el servidor genómico UCSC.

En la página UCSC Genome comprobamos los fragmentos del proyecto, que estén ubicados correctamente dentro del cromosoma 12. Esta localización se confirma en la figura 5.

2.3 Mutaciones en la proteína KRAS.

Este trabajo se centra en la mutación en el gen KRAS localizada en el exón 2, codón 12 del cromosoma 12. En la mutación G12D, GGT>GAT, la literatura establece que esta mutación genética es la responsable de la agresividad biológica del carcinoma de páncreas. La mutación G12D Aspartato otorga al carcinoma una mayor expansión local y una facilidad metastásica a distancia. También, la mutación G12V Valina en la misma localización, se asocia a un mal pronóstico de la enfermedad. Al igual que G12D, se encuentra en un número elevado de pacientes oncológicos (I. Roa et al 2013, S. Sho et al 2017). Las personas que poseen esta mutación genética, ya se considera que tienen una preneoplasia o un cáncer en estadio temprano.

El oncogén KRAS imposibilita la actuación de los supresores tumorales como son TP53, CDK2A y SMAD4 favoreciendo la actividad mitogénica descontrolada y supervivencia celular. Estas disposiciones son las que benefician la expansión tumoral y el mal pronóstico de la enfermedad.

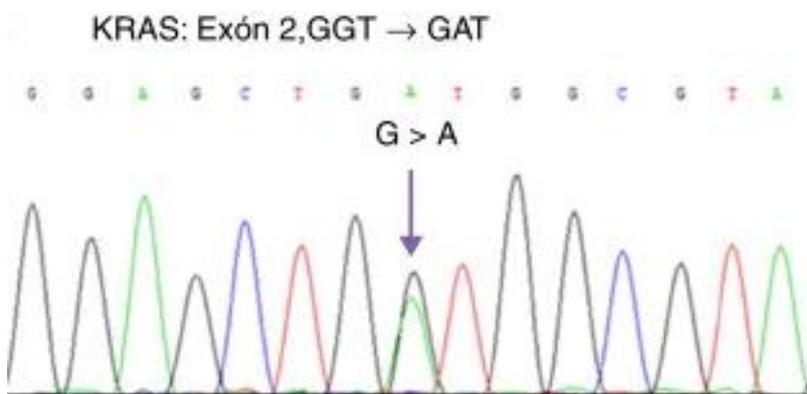


Figura 6. Mutación G12D, correspondiente con la localización 228 tomando como referencia la secuencia RefSeq NM004985.4

2.4 Análisis del gen KRAS.

Se hace el estudio de las mutaciones en los fragmentos del gen KRAS numerándolas según la secuencia consenso RefSeq, NM_004985.4.

La primera acción, será buscar como se ha comentado antes, las secuencias pancreáticas y no pancreáticas de la región de interés en los servidores genómicos. Las secuencias serán guardadas en formato fasta.

A continuación, se hará un alineamiento bastante complejo, puesto que las secuencias tienen tamaños muy diversos y algunas no llegan a cubrir las zonas de interés. Esta es la causa por la cual se divide el estudio en 5 fragmentos, para poder optimar al máximo el uso de las secuencias encontradas. Entre ellas, encontramos 26 pancreáticas y 69 no pancreáticas.

Para el alineamiento y edición, nos ayudamos con programas como MUSCLE, ClusterW Multiple Aligment y Bioconductor en R, aunque el ajuste termina manualmente.

2.5 Fragmentación de las secuencias de ADN.

Lo siguiente, es visualizar las secuencias en formato fasta en Mega5 o BioEdit, donde se pasa un control de calidad y se visualizan los cambios que se aprecian en las secuencias respecto a la secuencia RefSeq.

Fragmento1: Figura 7. Comprende la región entre el 193 – 292 con 48 secuencias coincidentes alineadas en la zona, que contienen 100pb. La posición de la mutación que aparece en el cáncer de páncreas, según la bibliografía G12D/G12V, corresponde a la posición 228 de la RefSeq. Teniendo la secuencia consenso GGAGCTG**G**TGGC, se produce una sustitución de Guanidina por Timina GGAGCTG**T**TGGC en la secuencia Sec79, que nombramos como G228T. Se trata de la única mutación encontrada en esa posición. En la Sec79 también hallamos la mutación G214C.

Podemos ver una sustitución G227T que se produce en las secuencias Sec49, Sec80, Sec83 y Sec85, al igual que la mutación C226G y G228C en las secuencias Sec82 y Sec83. Llama la atención la Sec43 y Sec88 que comparten varias mutaciones como T220C,

G236A, G241A, C246T y G284C. Además, la mutación G230A que solo está presente en la Sec43. También observamos una mutación A286T en la Sec92 de manera aislada. Igualmente, se halla una delección en la posición 204 y 205 de la Sec11.



Figura 7. Fragmento 1 visionado en BioEdit.

Fragmento2: Figura 8. Contiene 32 secuencias con 101pb comprendidas entre la posición 342 y 443 de la RefSeq. Siguiendo con el análisis se aprecia una sustitución A376T en las Sec11, Sec12, Sec13, Sec14, Sec82 y Sec93 y en la misma posición en la Sec8, la mutación A376C.

Se observa en las secuencias Sec43 y Sec88 que comparten varias mutaciones a lo largo de toda la región, como A361T, G368A, G379A, G387A, y una delección en la posición 396 y en la 420. Llama la atención que ambas tienen una inserción en la posición 416, la Sec43 tiene seis Guanidinas y la Sec88 tiene 2 Guanidinas, una Citosina y 8 Guanidinas más.

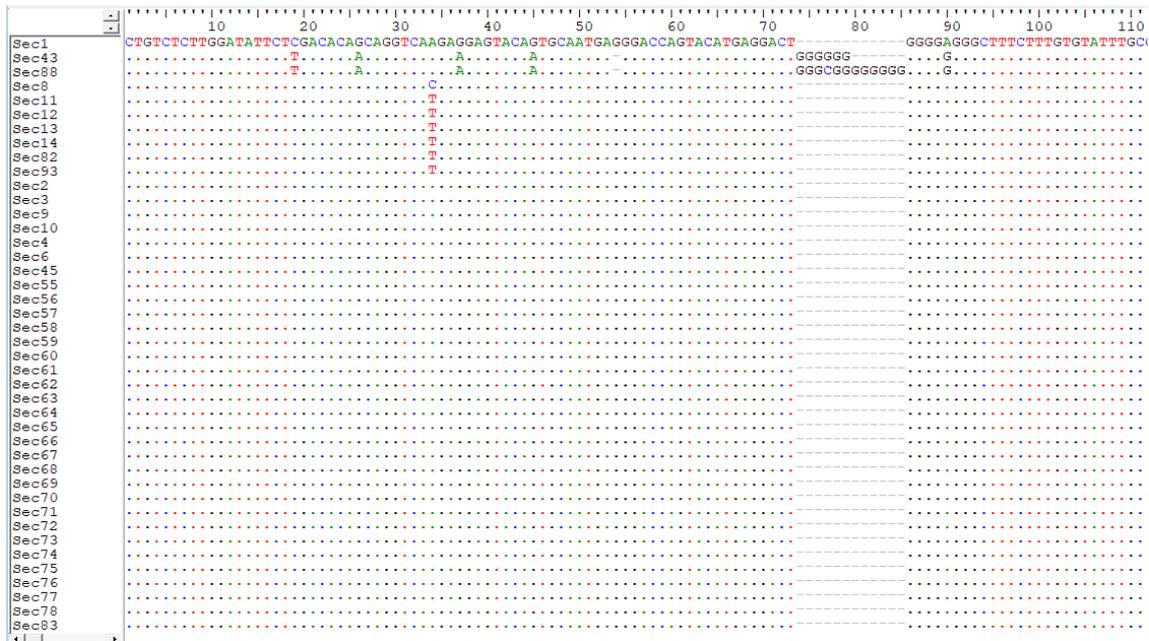


Figura 8. Fragmento 2 visionado en BioEdit.

Fragmento3: Figura 9. Esta comprendido entre la secuencia 303 y la 408, conteniendo 24 secuencias con una longitud de 178pb. En este fragmento está incluido el fragmento 2, donde se aumentan el número de pares de bases para ampliar la zona de estudio, pero disminuyen el número de secuencias que abarcan la zona. Se observan mutaciones a parte de las mencionadas en el fragmento 2.

En las Sec43 y Sec88 hay cuatro mutaciones más, que son A305G, C453T, C472G y C476A, unida a la misma delección en 387. También, la inserción de la Sec43 tiene seis G y la Sec88 tiene 2 G, una C y 8 G más en la posición 416.

Volvemos a encontrarnos la sustitución A376T en las Sec11, Sec12, Sec13, Sec14, Sec82 y Sec93 y en la misma posición la mutación A376C en la Sec8.



Figura 9. Fragmento 3 visionado en BioEdit.

Fragmento4: Figura 10. Comprendido entre la posición 193 hasta la 634 de la secuencia RefSeq, formado por 19 secuencias y con una longitud de 442pb. Las Sec43 y 88 tienen múltiples mutaciones a lo largo de toda la región añadidas a las de los 3 fragmentos anteriores, T213C, A420G, C453T, T472G, C476A, G485A, G500T, A520G, A575C, C578A, A598G, T604G, T613C, además de deleciones en 570-571 y 625-626.

En la Sec82 y Sec83 encontramos las sustituciones C266G, G268C y G620C.

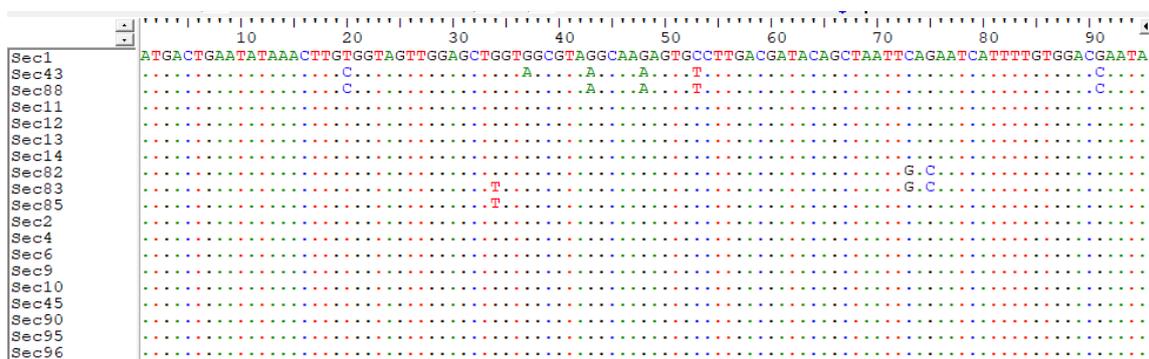


Figura 10. Fragmento 4 visionado en BioEdit.

Fragmento5: Figura 11. Comprendido entre las posiciones 243-301 con 49 secuencias, con 60pb. En este fragmento están incluidas todas las muestras pertenecientes a pacientes con carcinoma de páncreas. Las muestras pancreáticas son de la Sec15 a la Sec40, un total de 26, las demás son de varios tumores diferentes. Además de los cambios comentados en los anteriores fragmentos, también encontramos:

- Sec7: Mutación A294G.
- Sec15: Delección 258D.
- Sec16: Mutaciones T261C, A262C, A294T y la delección 258D.
- Sec17: Mutaciones A262C, G284C, A299T y la delección 258D.
- Sec18: Mutaciones T261C, A262C, A294T.
- Sec19: Mutaciones A251G, G253A, A256T, T261C, A262C, G284C, A296G, A299C, 300I, C302T, y la delección 258D.
- Sec20: Mutaciones 249D, A251G, A262C, A263T, A267G, A273T, A298C, A299T y la delección 258D.

- Sec21: Mutaciones A258G, T261C, A262C, A267G, A296G, A299T y la delección 256D.
- Sec22: Mutaciones A256C, C257G, A262C, A267G, A295G, A296C, A298C, A299T, 300D, 301G, y la delección 258D.
- Sec24: Mutaciones A262C, A299T y la delección 258D.
- Sec25: Mutaciones G245C, T248G, A251G, A256C, C257G, C260A, T261C, A262C, A263T, T264G, A270G, G284C, A296G, A298C, A299C, G302C y la delección 258D.
- Sec27: Mutaciones T261C, A262C, T264G, G284C, A296G, A301G y la delección 258D.
- Sec28: Mutaciones A258G, A262C, A267G, A299T y la delección 246D.
- Sec29: Mutaciones A258G, A262C, A299T y la delección 261D.
- Sec30: A258G, A262C, A267G y la delección 256D.
- Sec31: Mutaciones G253A, A256C, C257G, T260C, A262T, G284C, la inserción 300I y la delección 258D.
- Sec32: Mutación A273T y la delección 258D.
- Sec37: Mutaciones A262C, A299T y la delección 258D.
- Sec39: Mutaciones T261T, A262C y la delección 258D.
- Sec41: Mutación A284C.

Vemos como la mayoría de las mutaciones se producen en secuencias pancreáticas, a excepción de la Sec41 que pertenece a vesícula biliar y la Sec7 a colon y pulmón. Además, observamos 7 muestras pancreáticas que no tienen ninguna mutación en el fragmento.

Llaman la atención cambios similares que se producen en algunas secuencias:

A251G: Sec19, Sec20 y Sec25.

C253A: Sec19 y Sec31.

A256T: Sec19, Sec21(256D), Sec22(A256C), Sec25(A256C), Sec30(256D) y Sec31 (A256C).

258D: Sec15, Sec16, Sec17, Sec19, Sec20, Sec21 (A258G), Sec22, Sec24, Sec25, Sec27, Sec28 (A258G), Sec29 (A258G), Sec30, Sec31, Sec32, Sec37 y Sec39.

T261C: Sec16, Sec18, Sec19, Sec21, Sec25, Sec27, Sec29 (261D) y Sec39.

A262C: Sec16, Sec17, Sec18, Sec19, Sec20, Sec21, Sec22, Sec24, Sec25, Sec27, Sec28, Sec29, Sec30, Sec31 (A262T), Sec37 y Sec39.

A267G: Sec20, Sec21, Sec20, Sec28 y Sec30.

A273T: Sec20 y Sec32.

G284C: Sec17, Sec19, Sec25, Sec27, Sec31 y Sec41.

A294T: Sec7 (A294G), Se16 y Sec18.

A296G: Sec19, Sec21, Sec22, Sec25, Sec27

A298C: Sec20, Sec22 y Sec25.

A299T: Sec17, Sec19(A299C), Sec20, Sec21, Sec22, Sec24, Sec25(A299C), Sec28, Sec29 y Sec37.

Se observan cambios destacables, en la posición 258 encontramos 17 secuencias con cambios. En la 262, se hallan 16 y en la 299 otros 10.



Figura 11. Fragmento 5 visualizado en BioEdit.

Podemos visualizar en la figura 11, el alineamiento de las secuencias del fragmento 5 donde se incluyen las pancreáticas. Entre la zona 254-260 con referencia a la RefSeq, observamos una zona con muchas mutaciones, deleciones e inserciones. Quizá se puedan deber a un mal estado de las secuencias publicadas en el GenBank o quizá esos

cambios puedan tener mucha significancia en el carcinoma de páncreas. Sería interesante un posterior estudio sobre las mutaciones que padecen estos pacientes.

Como conclusión, sobre los cambios producidos en los cinco fragmentos, se pueden observar en la tabla del Anexo 2, donde se han anotado todos los cambios de las secuencias. Podemos apreciar como algunos cambios se producen reiteradamente en algunas secuencias.

Encontramos como más destacables, la mutación A376T que aparece en 7 secuencias, la A299T que aparece en 10, A262C en 17 secuencias, T261C en 10, A258G en 18 y 256D en 7, las que más se repiten.

2.6 Análisis de los cambios genéticos.

En la tabla 5 se resume el número de secuencias encontradas y las mutaciones que tienen según el diagnóstico de cáncer.

	Secuencias	Mutaciones
Pacientes con carcinoma pancreáticos sin mutaciones genéticas.	7	0
Pacientes con carcinoma pancreático con mutaciones genéticas.	19	107
Pacientes con carcinoma no pancreático sin mutaciones genéticas	51	0
Pacientes con carcinomas no pancreáticos con mutaciones genéticas.	18	84

TABLA5. Resumen del número de secuencias y mutaciones en pacientes con carcinomas pancreáticos y no pancreáticos.

Apoyándonos en el Anexo 1 y 2, se hace una valoración de las mutaciones genómicas encontradas en las secuencias de ADN descargadas de las bases de datos genéticas.

Encontramos que las secuencias que más mutaciones acumulan con las Sec43 y Sec88 que corresponden a secuencias de carcinomas de origen desconocido, con 29 mutaciones. Aunque tengan el mismo número de mutaciones no coinciden totalmente. Se diferencian en que la Sec88 tiene la mutación G620C, mientras que la Se43 tiene la mutación G230A. Ambas secuencias abarcan todos los fragmentos del proyecto. Tienen el porcentaje mayor de mutaciones con un 6.56%.

Las muestras pertenecientes a pacientes con carcinoma de páncreas destacan, la mayoría, por tener un alto porcentaje de mutaciones en el fragmento número 5, siendo el único fragmento en el que participan. De las 26 muestras que, 19 contienen mutaciones, alcanzando un porcentaje de 21.7% en el caso de la Sec25, o un 18,3% en las Sec19 y Sec22. Por otro lado, 7 de las secuencias pancreáticas no contienen ninguna mutación.

Los cambios que más se han hallado, son la delección en la posición 258, respecto a la secuencia RefSeq, que se aprecia en 15 secuencias y el cambio de Adenina por Guanidina que se da en otras 4 secuencias en esa posición. Igualmente, en la posición 261 se producen 9 cambios de Timina por Guanidina y una delección. En la siguiente posición, 262, se producen 17 cambios de Adenina por Citosina. En la posición 299 la Adenina es sustituida en dos secuencias por Citosina y en siete por Timina. Todos estos cambios, corresponden a pacientes con carcinoma de páncreas.

La mutación G284C se produce en 8 secuencias, 7 de ellas en carcinoma de páncreas y una en carcinoma de origen desconocido. En la posición 376 se producen seis cambios de Adenina por Citosina y uno por Timina en dos pacientes con carcinoma de pulmón, cerebro, colon y de origen desconocido.

	Número de muestras	Positivas	% Positivas	Número de Mutaciones	Media número pb	Media de pb del estudio
Pancreáticas	26	19	73,07	107	152,38	61,84
No Pancreáticas	69	18	28.57	84	408485,04	213,68

TABLA 6. Tabla de información sobre las muestras que poseen las secuencias de ADN y su longitud.

En la tabla 6 se analizan los tantos por cierto de las muestras de carcinomas pancreáticos y no pancreáticos. En este caso, hablamos de positividad, para referirnos a las secuencias que contienen mutaciones. Apreciándose que el tanto por ciento en mutaciones es muy superior en muestras de carcinoma de páncreas, con un 73.07%, mientras que las no pancreáticas poseen un 28.57%.

A pesar de haber bastantes más secuencias no pancreáticas con 69 en total, y tener unas secuencias de mayor longitud con una media de 213,68pb y están incluidas en más fragmentos, hallamos que los pacientes pancreáticos tienen más mutaciones. Encontramos un total de 107 mutaciones en secuencias más cortas con 61.84pb de media en estos 26 pacientes que solo se incluyen en el fragmento 5. A diferencia de los no pancreáticos con 84 mutaciones en total.

Según las mutaciones acumuladas por ambas poblaciones, los pancreáticos y los no pancreáticos, se puede afirmar que acumulan más mutaciones los pacientes con carcinomas pancreáticos. Quizá este acúmulo de mutaciones tenga una potencial influencia en el padecimiento de la enfermedad según los resultados obtenidos.

2.7 Análisis descriptivo.

Se realiza un análisis descriptivo para estudiar los datos que se han obtenido en los servidores genómicos. Es interesante hacer un estudio de las mutaciones encontradas en las secuencias de ADN de los pacientes y hacer una relación con los tipos de cáncer que padecen. Como consecuencia, poder establecer una relación entre el tipo de cáncer y las mutaciones, que pueden reflejar la influencia de las mutaciones en el padecimiento de la enfermedad.

Para realizar el análisis, las muestras se han separado en dos grupos. El primero son las secuencias pancreáticas y el segundo las no pancreáticas representadas en la tabla 7. En ella, se especifican la longitud de las secuencias, los cambios genéticos de cada secuencia y el tanto por cierto de mutaciones con respecto a su longitud de las secuencias para compararlas.

Sec	Mut	pbGenBank	pb	Mut%	Tipo_Cancer
Sec15	1	161	60	1,67	Pancreatic
Sec16	4	162	60	6,67	Pancreatic
Sec17	4	163	60	6,67	Pancreatic
Sec18	3	166	60	5	Pancreatic
Sec19	11	171	60	18,3	Pancreatic
Sec20	9	160	60	15	Pancreatic
Sec21	7	163	60	11,7	Pancreatic
Sec22	11	156	60	18,3	Pancreatic
Sec24	3	131	60	5	Pancreatic
Sec25	13	184	60	21,7	Pancreatic
Sec27	7	162	60	11,7	Pancreatic
Sec28	5	157	60	8,33	Pancreatic
Sec29	4	165	60	6,67	Pancreatic
Sec30	4	134	60	6,67	Pancreatic
Sec31	8	173	60	13,3	Pancreatic
Sec32	2	153	60	3,33	Pancreatic
Sec33	5	163	60	8,33	Pancreatic
Sec37	3	130	60	5	Pancreatic
Sec39	3	127	60	5	Pancreatic
Sec7	1	390	108	0,93	No pancreatic
Sec8	1	4418	287	0,35	No pancreatic
Sec11	3	1693	442	0,68	No pancreatic
Sec12	1	567	442	0,23	No pancreatic
Sec13	1	607	442	0,23	No pancreatic
Sec14	1	607	442	0,23	No pancreatic
Sec39	3	127	60	5	No pancreatic
Sec41	1	123	108	0,93	No pancreatic
Sec43	29	865	442	6,56	No pancreatic
Sec49	1	136	108	0,93	No pancreatic
Sec79	2	113	108	1,85	No pancreatic
Sec80	1	113	108	0,93	No pancreatic
Sec82	5	450	442	1,13	No pancreatic
Sec83	5	450	442	1,13	No pancreatic
Sec85	1	5775	442	0,23	No pancreatic
Sec88	29	5458	442	6,56	No pancreatic
Sec92	1	111	108	0,93	No pancreatic
Sec93	1	560	178	0,56	No pancreatic

TABLA 7. Representación de las muestras pancreáticas y no pancreáticas, con el número de mutaciones que contiene cada secuencia.

Se comienza un estudio descriptivo en R, haciendo un preprocesado de datos. En un boxplot, hacemos una primera visualización de los datos que se han recopilado.

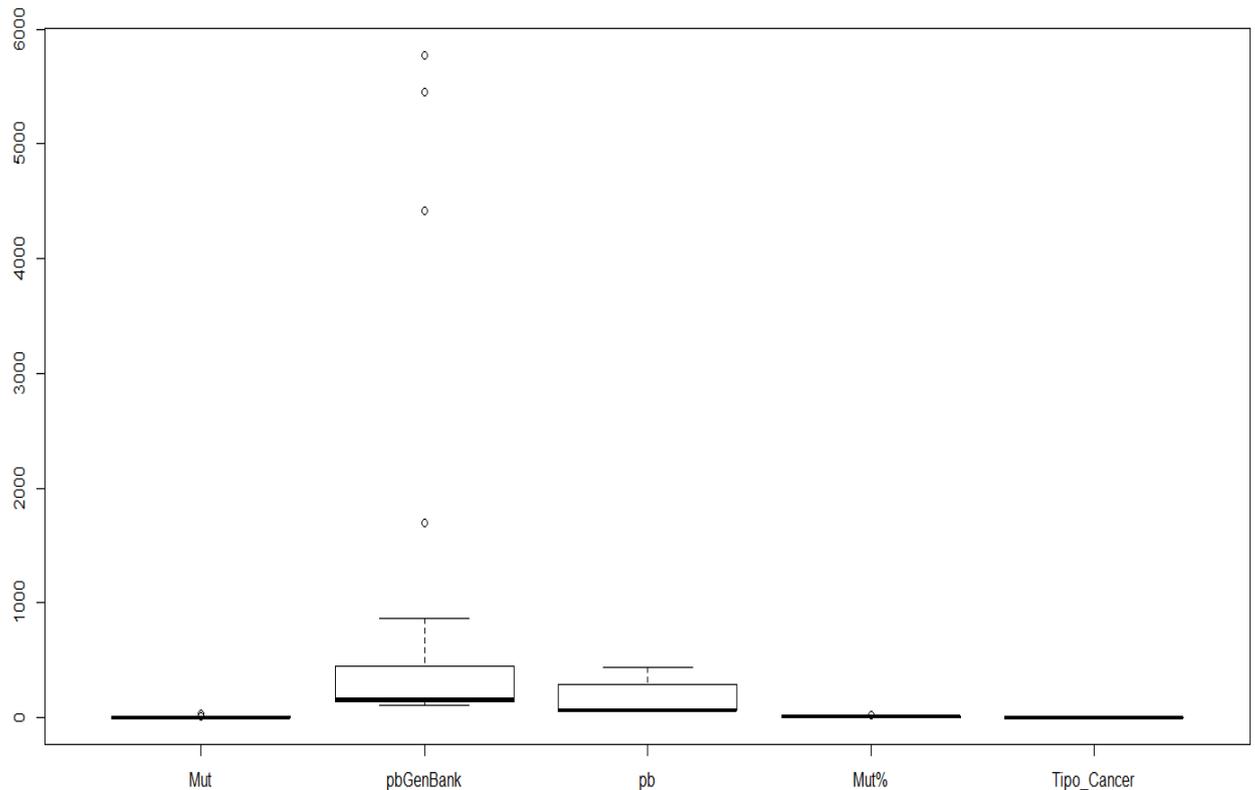


Figura 12. Representación gráfica de los datos recabados. En el eje y se cuantifican el número de mutaciones, de pares de bases encontradas en el GenBank, las pares de bases que se usan el proyecto, el tanto por ciento de las mutaciones según la longitud de la secuencia y el tipo de cáncer al que corresponden.

En la figura 12, vemos los datos que se han podido recabar en las bases de datos genómicas. Lo más destacable en la primera visualización de los datos, es la longitud de las secuencias encontradas en el GenBank, que varía mucho de unas secuencias a otras. Pero en cambio, el uso máximo de la longitud de las secuencias en el proyecto es de 442pb.

Se comienza el análisis descriptivo de los datos que hemos recabado hasta el momento. Empezamos el análisis con el paquete *readxl*, visualizando nuestros datos divididos en dos grupos, pancreáticos y no pancreáticos.

En primer lugar, se realiza un *summary()* observacional.

Sec	Mut	pb
Length:37	Min. : 1.000	Min. : 60
Class :character	1st Qu.: 1.000	1st Qu.: 60
Mode :character	Median : 3.000	Median : 60
	Mean : 5.243	Mean :170
	3rd Qu.: 5.000	3rd Qu.:287
	Max. :29.000	Max. :442
pbGenBank	Mut%	Tipo_Cancer
Min. : 111.0	Min. : 0.2262	Min. :0.0000
1st Qu.: 136.0	1st Qu.: 0.9259	1st Qu.:0.0000
Median : 163.0	Median : 5.0000	Median :0.0000
Mean : 690.4	Mean : 5.6133	Mean :0.4865
3rd Qu.: 450.0	3rd Qu.: 6.6667	3rd Qu.:1.0000
Max. :5775.0	Max. :21.6667	Max. :1.0000

Figura 13. *summary()* de los datos en R.

Como datos de interés en la figura 13, podemos apreciar que la media de las mutaciones es de 5.24 con un mínimo de 1 y un máximo de 29. En cambio, en el tanto por cierto de mutaciones, es un 5.61, con un mínimo de 0.22 y un máximo de 21.66. La media de pares de bases usadas en el estudio es de 170 en los cinco fragmentos.

Se relaciona el número de mutaciones encontradas en los dos grupos, los pacientes con carcinoma de páncreas y no pancreáticos. Utilizando el paquete *psych* y la función *describeBy()* para hacer una estadística descriptiva por grupos. En la figura 14 se observa que, en el grupo de los pacientes con carcinoma pancreático, tienen una media superior a los no pancreáticos.

```

Descriptive statistics by group

group: 0
vars  n mean  sd median trimmed  mad min max range skew kurtosis  se
x1    1 19  5.63 3.39  4    5.47 1.48  1 13   12 0.72  0.75 0.78
-----

group: 1
vars  n mean  sd median trimmed  mad min max range skew kurtosis  se
x1    1 18  4.83 8.89  1    3.56 0    1 29   28 2.18  3.08 2.1

```

Figura 14. Resultados obtenidos con la función *describeBy()* en R.

Pancreáticos con una media de 5.63 ∓ 3.39 de desviación estándar.

No pancreáticos con una media de 4.83 ∓ 8.89 de desviación estándar.

Realizamos el mismo estudio pero con el tanto por ciento de mutaciones según la longitud de la secuencia con la función *describeBy()*.

```

Descriptive statistics by group

group: 0
vars  n mean  sd median trimmed  mad  min  max range skew kurtosis  se
x1    1 19  9.39 5.64  6.67 9.12 2.47 1.67 21.67  20 0.72  -0.75 1.29
-----
--
group: 1
vars  n mean  sd median trimmed  mad  min  max range skew kurtosis  se
x1    1 18  1.63 2.09  0.93  1.41 0.7 0.23 6.56  6.33 1.58 0.84 0.49

```

Figura 15. Resultados obtenidos con la función *describeBy()*.

En la figura 15 encontramos que los resultados varían.

Pancreáticos con una media de 9.39 ∓ 5.64 de desviación estándar.

No pancreáticos con una media de 1.63 ∓ 2.09 de desviación estándar.

Se incrementa la desviación estándar en los pancreáticos y disminuye entre los no pancreáticos, al igual que la media, en los pancreáticos sube de 5.63 a 9.39 y en los no pancreáticos baja de 4.83 a 1.63.

Lo siguiente será visualizar los datos y ver si encontramos algún dato atípico con la función *boxplot()* dentro del número de mutaciones.

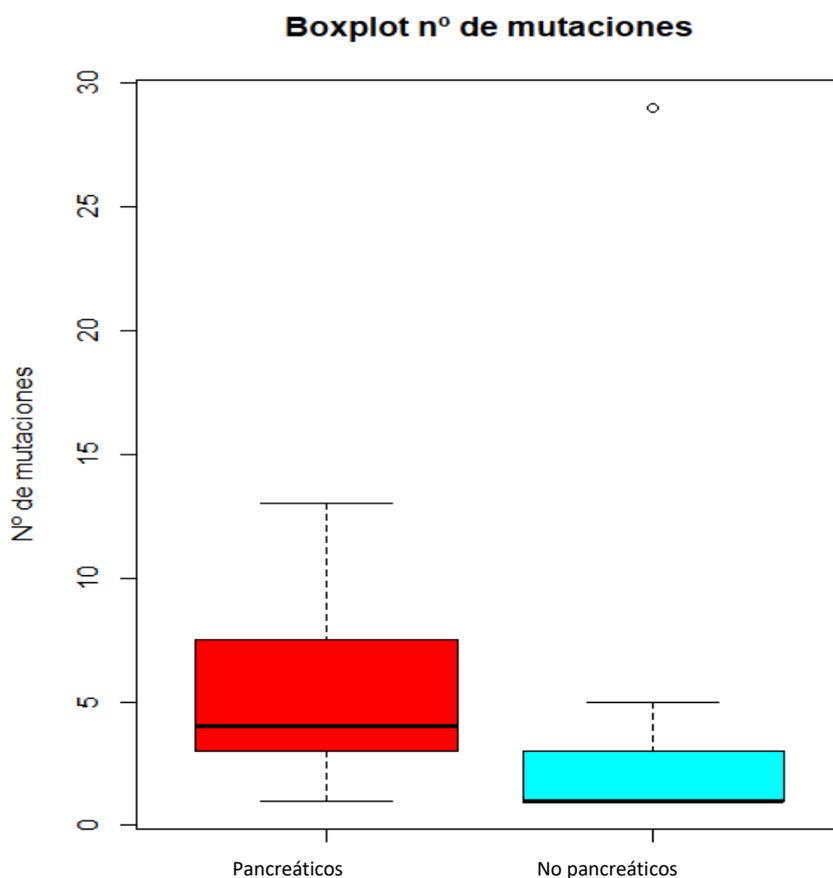


Figura 16. Boxplot del número de mutaciones que contienen las secuencias de pacientes con carcinoma de páncreas comparada con carcinomas de otro tipo de carcinomas.

Podemos observar que la gráfica de la figura 16 es asimétrica y un dato atípico en los pacientes no pancreáticos. El número de mutaciones es mayor en los pacientes pancreáticos, pero a la vez, en la gráfica se observa más dispersión.

Realizamos el mismo gráfico, pero con el tanto por ciento de mutación acumuladas, según la longitud de secuencia usada en los fragmentos que realizamos anteriormente.

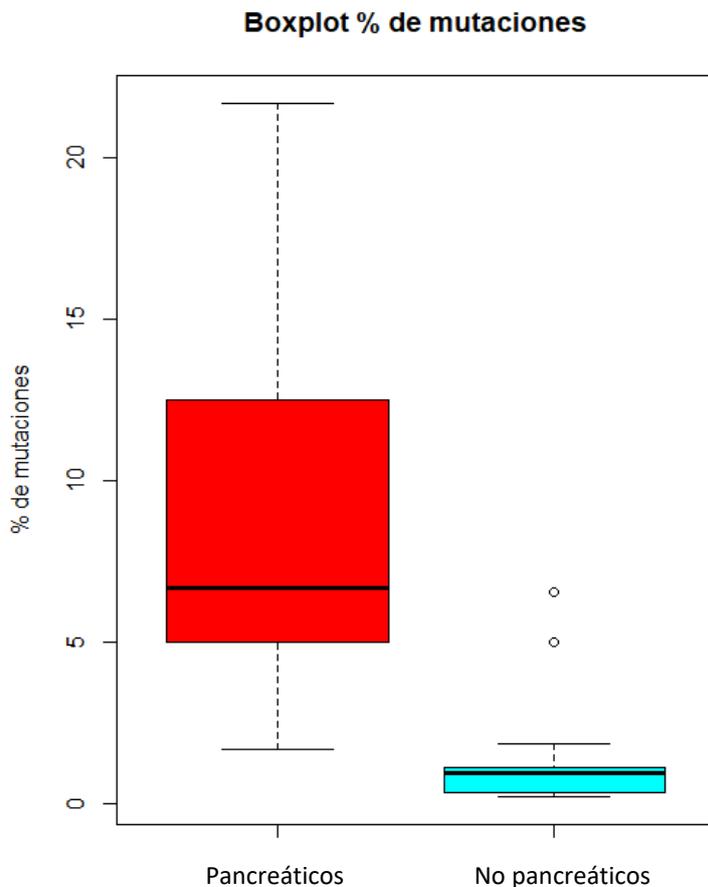


Figura 17. Boxplot con el tanto por ciento de mutaciones con respecto al tamaño de la secuencia que contienen las secuencias de pacientes con carcinoma de páncreas comparada con carcinomas de otro tipo.

En la figura 17 en esta ocasión, vemos que la media de mutación también es mayor en los pacientes pancreáticos, al igual que en la gráfica anterior. A la vez, el grado de dispersión en los pancreáticos es mayor. En los pacientes no pancreáticos vemos dos datos atípicos.

En ambas gráficas, tanto en la figura 16 como en la figura 17, se pone de manifiesto que el número de cambios en las secuencias de ADN, es mayor en los pacientes con carcinoma de páncreas.

Seguimos comprobando los datos del proyecto a través del test de normalidad de Shapiro-Wilk, obteniendo un p-value de 4.903e-07, inferior a 0.05. Con este dato,

podemos decir que siguen una distribución normal, cumpliendo el supuesto de normalidad y siendo significativos.

Se prosigue con una prueba de ANOVA, o análisis de la varianza, para comprobar si las medias de ambas poblaciones son iguales. La hipótesis nula dirá que las medias de la población son iguales, mientras que la alternativa dirá que ambas poblaciones, las de pacientes con carcinoma de páncreas y pacientes con otro tipo de tumores, son diferentes. Realizamos la prueba con la función *aov()*.

El resultado de la prueba de ANOVA realizado sobre el número de mutaciones, nos dice que $F(1,35)=0.718$, donde se hace notorio que no existen diferencias entre los dos grupos. No son estadísticamente significativas. Es decir, el número de mutaciones que padecen en enfermos pancreáticos no son estadísticamente diferentes a las mutaciones en pacientes con otro tipo de cáncer.

En el caso de realizar la prueba ANOVA pero con el tanto por cierto de mutaciones obtenemos como resultado que $F(1,35)= 3.75e-06$. En este caso, si hay diferencias estadísticamente significativas, puesto que se tiene en cuenta la longitud de las secuencias usadas en el experimento. Ambas poblaciones, tanto pancreáticas como no pancreáticas, acumulan mutaciones de manera diferente. Se confirma los resultados que se visualizaban en los boxplot, donde se observaba un mayor acumulo de mutaciones en los pacientes con carcinoma de páncreas.

2.8 Estudio poblacional.

El último paso, es el estudio poblacional de las muestras. Hubiera sido de gran interés haber encontrado más información sobre las secuencias descargadas de los bancos de datos genómicos, ya que, no tenemos datos de algunas secuencias.

Con las alineaciones ya resueltas en los análisis anteriores, se procede a hacer un estudio filogenético de las secuencias de ADN. Con este análisis se quiere observar la lejanía o cercanía de las secuencias entre sí y la relación que existe entre ellas.

Los árboles filogenéticos se realizan con el programa Mega7 y Clustal de Omega, por el método de Neighbor-joining y Máxima verosimilitud (Maximum likelihood).

El árbol filogenético de Neighbor-joining, se basa en el criterio de mínima evolución en la distancia entre secuencias, minimizando la longitud del árbol. Se creó por Saitou & Nei en 1987 y es aditivo, pero no enraizado. Consiste en que, a través de un nodo interno, o secuencia, se coloca la siguiente secuencia con menor distancia y así sucesivamente. Considerando por distancia, cada cambio de aminoácido en la secuencia de ADN. Cada rama es indicativa de un cambio evolutivo y se construye a través de matrices de distancia.

El método estadístico de Máxima verosimilitud está basado en el modelo de Tamura-Nei. Para la construcción del árbol, se utilizan los algoritmos probabilísticos sobre un modelo de evolución. Se refleja la probabilidad de cada posición en la secuencia de ADN. Por consiguiente, se construyen todos los árboles posibles seleccionando el más verosímil.

En este proyecto se construyen árboles por ambos métodos, a partir de los cinco fragmentos que se crearon en el análisis anterior de secuencias.

Fragmento 1.

La primera representación será por el método de Neighbour-Joining en el fragmento 1, representación cladograma:

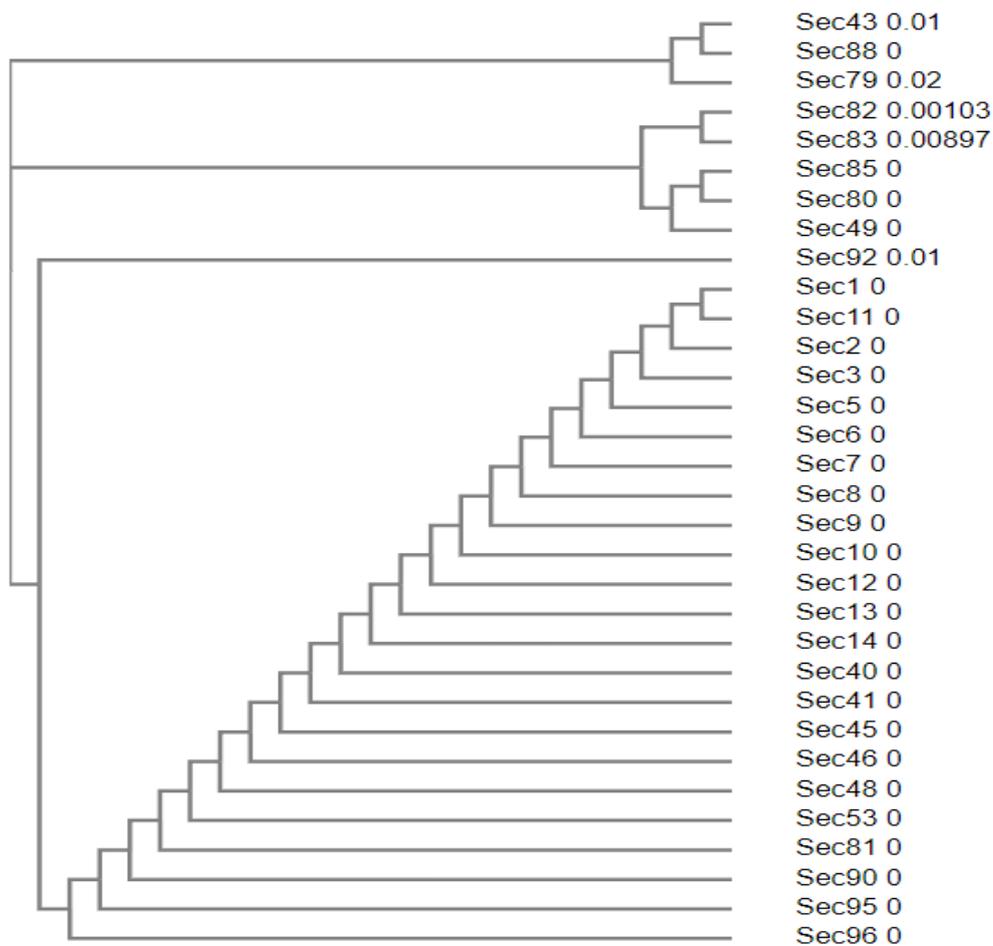


Figura 18. Árbol filogenético Neighbour-Joining del fragmento 1.

En la figura 18 se observa poca distancia en los brazos en el árbol real, porque existen pocos cambios en algunas secuencias, Se pone de manifiesto, que existen dos brazos del árbol que se desplazan del resto de muestras. El primero, contiene la Sec43, Sec88 y Sec79, siendo ésta última perteneciente a un osteosarcoma, mientras que las otras dos son de origen desconocido. El segundo brazo, incluye las Sec82 y Sec83. Dentro de la misma rama, apreciamos a Sec85, Sec80 y Sec49, todas ellas pertenecientes a carcinoma de pulmón, a excepción de la Sec85 que es de origen desconocido.

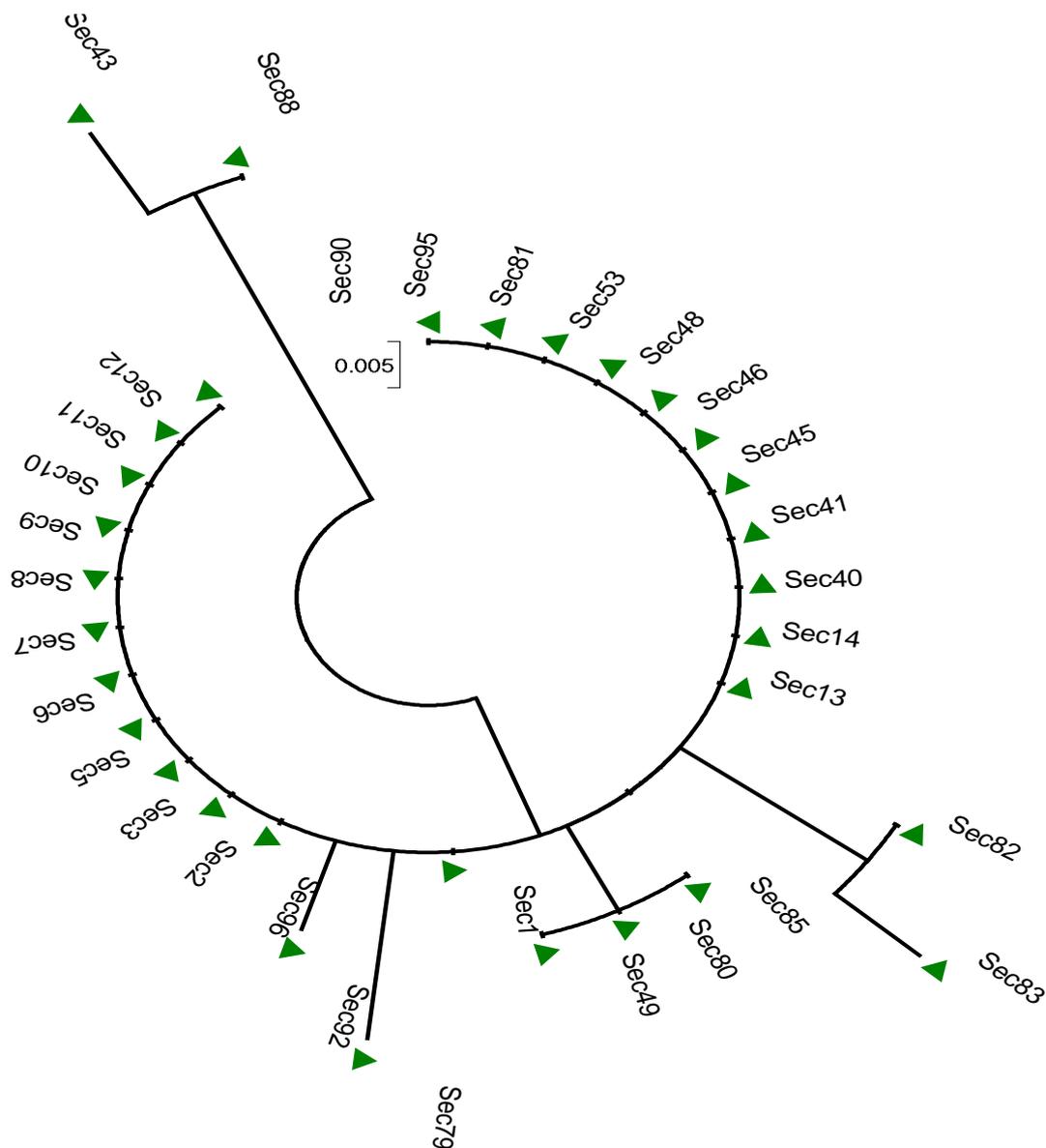


Figura 19. Árbol filogenético Maximum Likelihood del fragmento 1.

A través de Máxima verosimilitud en la figura 19, vemos como se vuelven a unir las muestras Sec82 y Sec83 de pulmón con una mayor distancia, debido a la acumulación del número de cambios en la secuencia. Las Sec43 y Sec88 de origen desconocido, se localizan en el brazo con mayor distancia por la cantidad de mutaciones acumuladas. Por otro lado, pero con menor distancia, la Sec85, Sec80 y Sec49, también de pulmón. Igualmente apreciamos el brazo largo de la Sec79 de osteosarcoma.

Fragmento 2

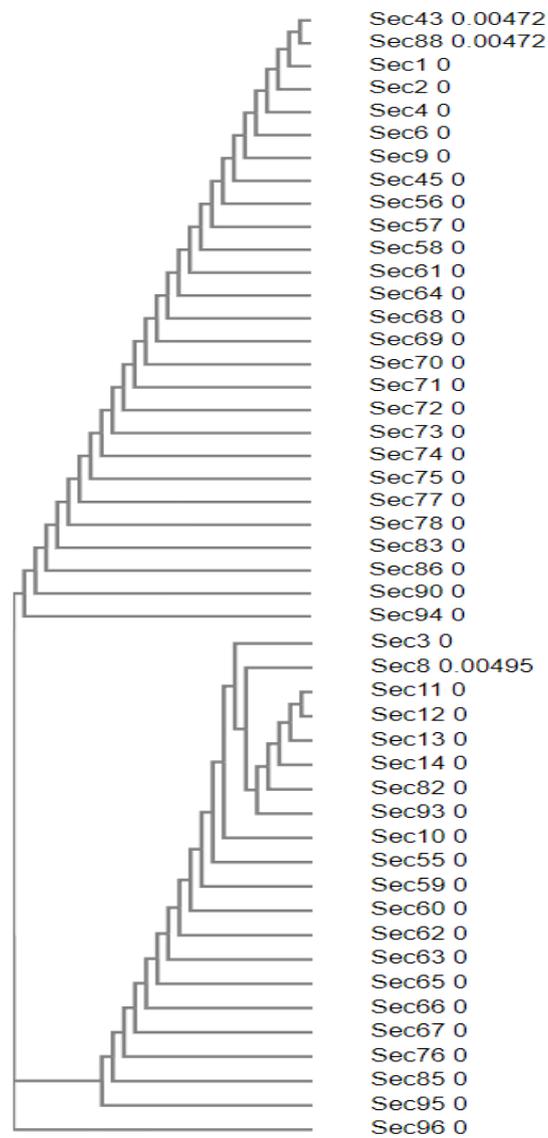


Figura 20. Árbol filogenético Neighbour-joining del fragmento 2.

En el árbol de la figura 20, parece que las muestras se dividen en dos grupos. El primero, con mucha similitud y el segundo en el que se ven más diferencias en el interior de sus ramas. La muestra Sec96 de carcinoma desconocido, se separa de las demás. Vemos en el mismo brazo, dentro de la segunda rama, la unión de las muestras Sec11 (cerebro), Sec12 (desconocido), Sec13 (desconocido), Sec62 (vejiga) y Sec93 (pulmón), como más característico.

Fragmento3

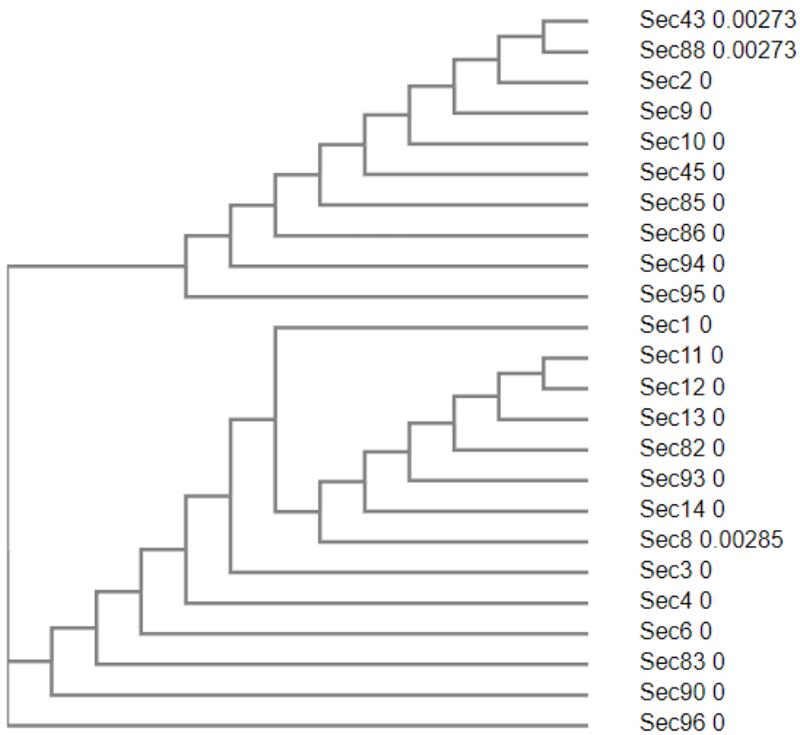


Figura 22. Árbol filogenético Neighbour-Joining del fragmento 3.

En este fragmento representado en la figura 22 intervienen menos secuencias. Observamos la formación de dos grupos y la separación total de la muestra Sec96 de origen desconocido. Volvemos a ver la agrupación de las muestras Sec43 y Sec88.

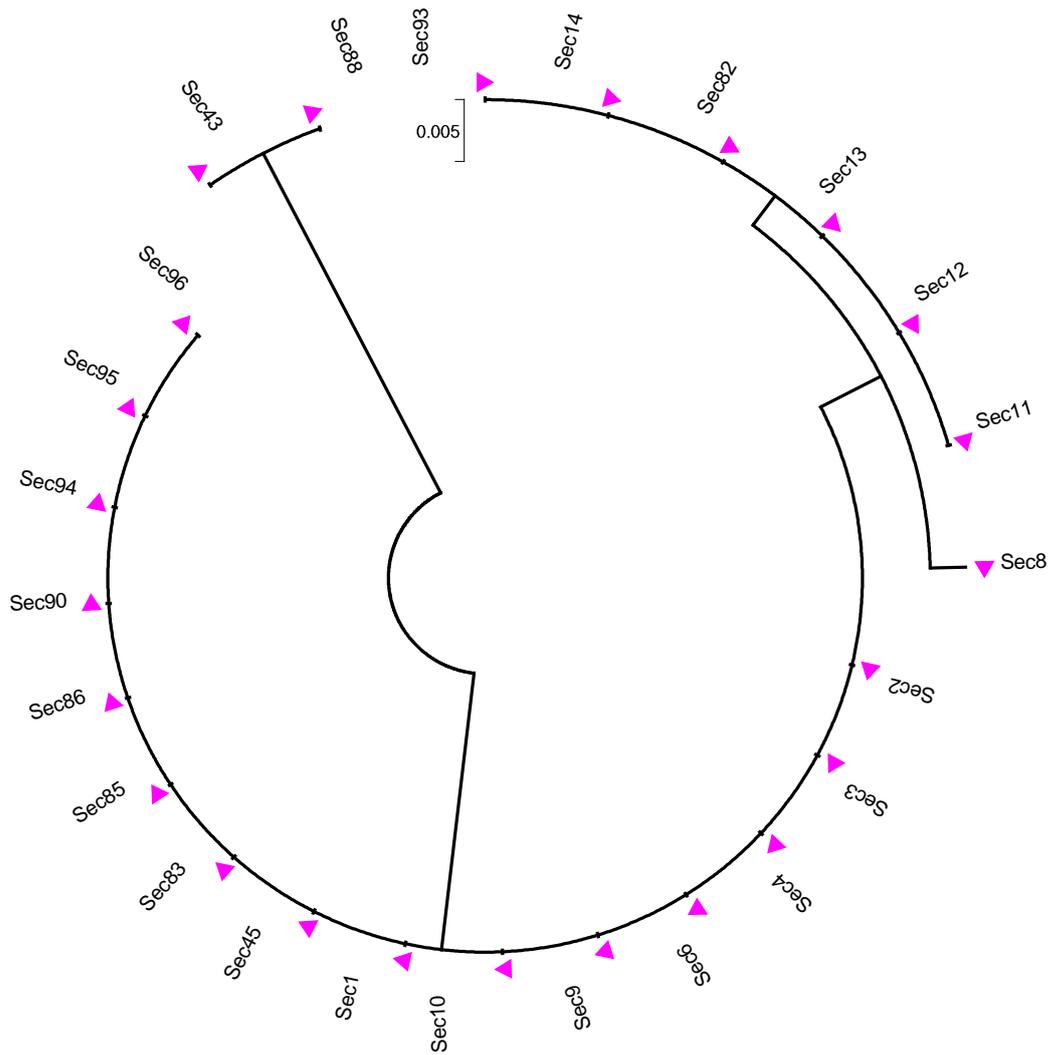


Figura 23. Árbol filogenético Maximum Likelihood del fragmento 3.

En la figura 23, las muestras Sec43 y Sec88 se vuelven a agrupar juntas, separadas de las demás muestras y además volvemos a ver la agrupación de Sec11, Sec12, Sec13, Sec14, Sec82, Sec93, como en el fragmento 2 por Máxima verosimilitud.

Fragmento 4.

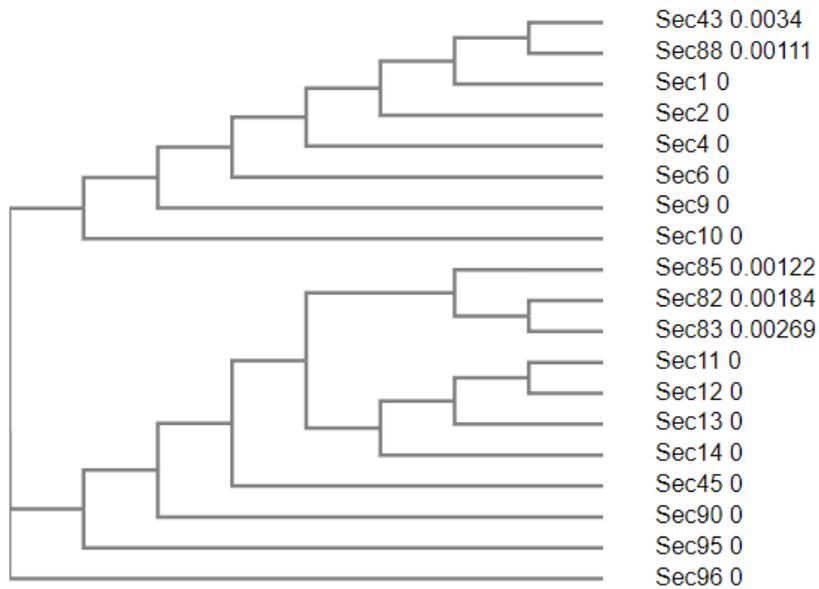


Figura 24. Árbol filogenético Neighbour-Joining del fragmento 4.

El fragmento 4, representado en la figura 24, es el mayor de todos con una longitud de 442pb, aunque incluye menos muestras, como ya mencionamos anteriormente. En la figura 24 apreciamos como la muestra Sec96 aparece separada del resto. Advertimos la formación de dos grupos, las Sec43 y Sec88 que permanecen juntas, pero dentro de la primera rama grande. En el segundo grupo, la Sec82 y Sec83 también unidas, pero en la misma rama que la secuencia Sec85.

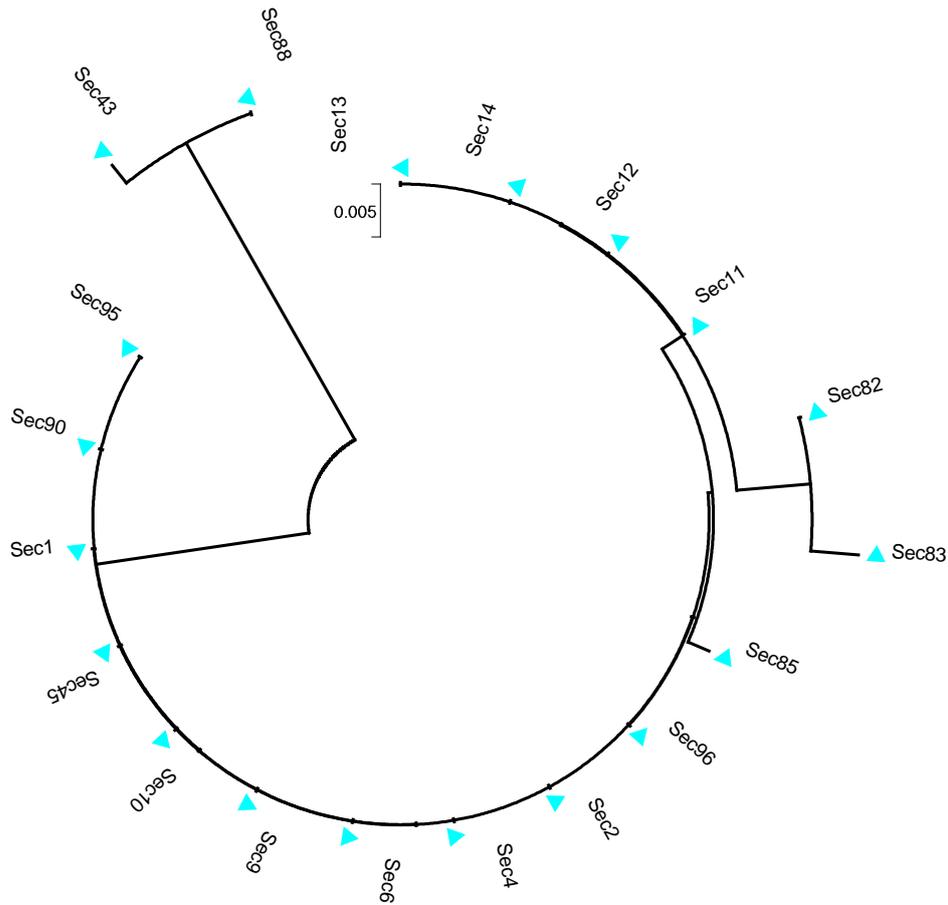


Figura 25. Árbol filogenético Maximum Likelihood del fragmento 4.

Por este segundo método, en la figura 25, destacamos que las muestras Sec43 y Sec88 aparecen agrupadas y fuera de las demás ramas. La Sec82 y Sec83 vuelven a unirse en la misma rama que la Sec11 a la Sec14 que tienen el mismo comportamiento, como datos más destacables.

Fragmento 5

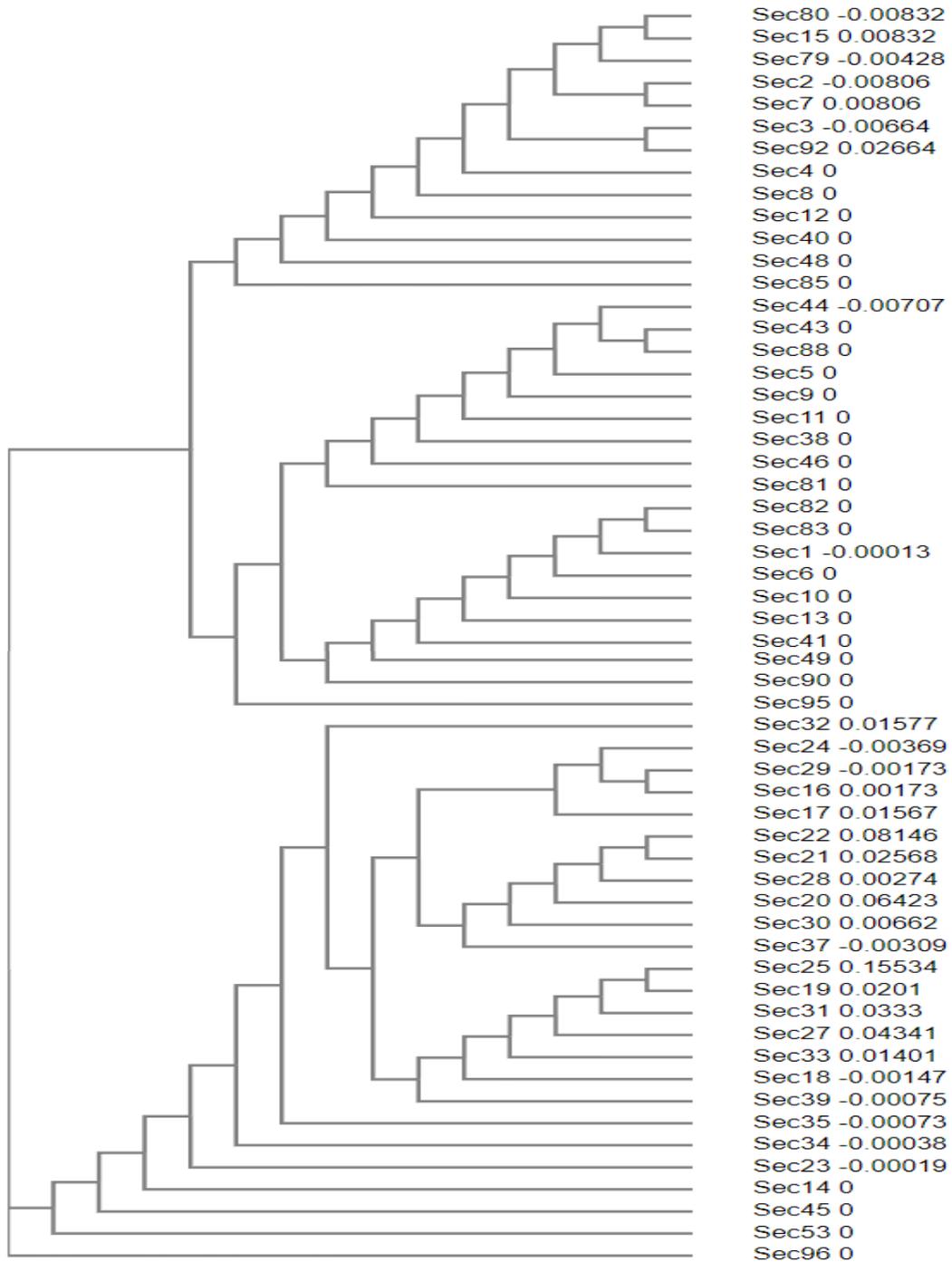


Figura 26. Árbol filogenético Neighbour-Joining del fragmento 5.

Este fragmento de la figura 26 es muy significativo, puesto que en él se incluyen todas las secuencias de carcinoma de páncreas, donde observamos como todas las secuencias que tienen mutaciones, se distribuyen por todo el árbol. Las muestras de pacientes pancreáticos sin mutaciones permanecen unidas en las mismas ramas.

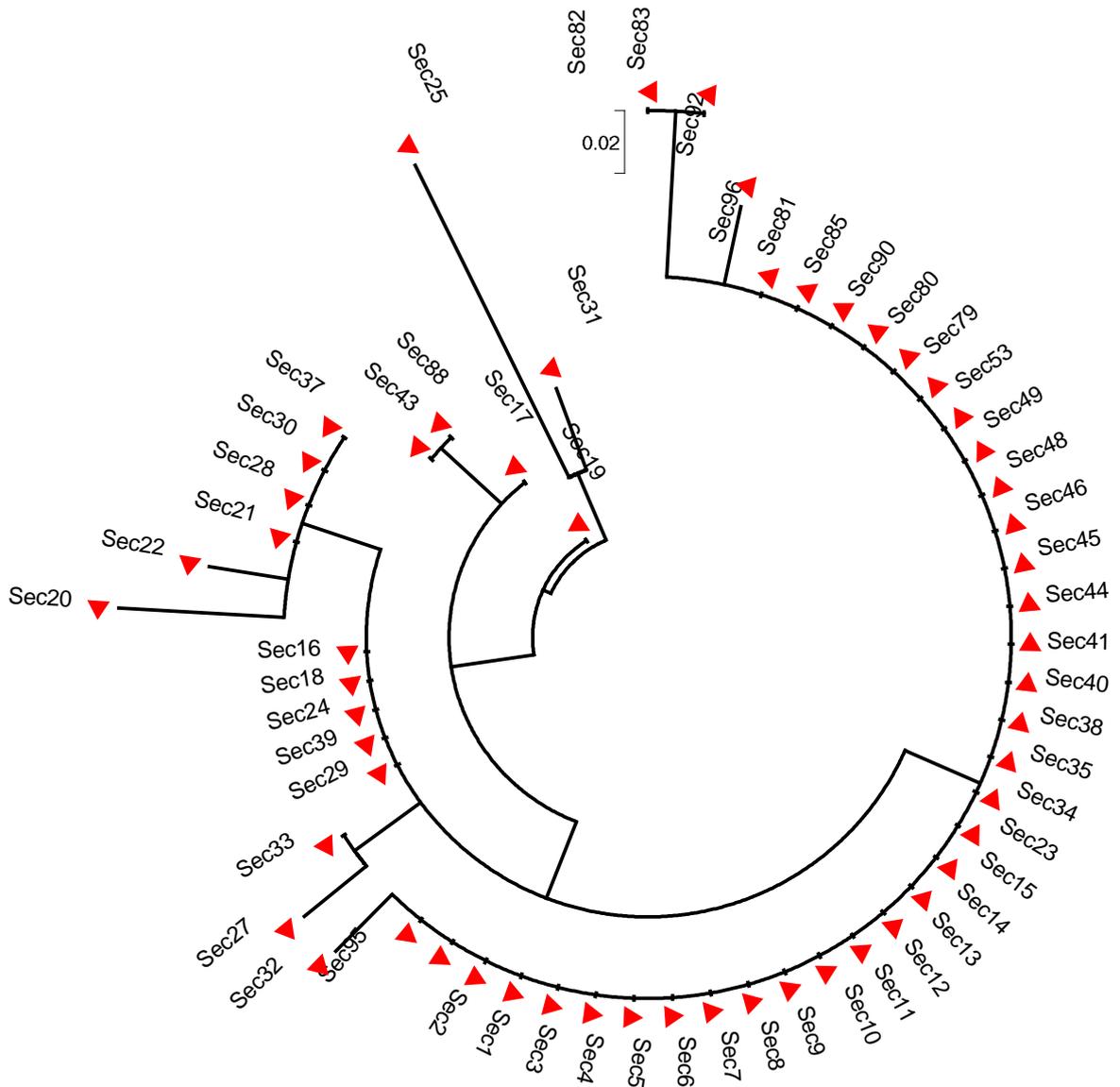


Figura 27. Árbol filogenético Maximum Likelihood del fragmento 5.

En la figura 27 se aprecia como la muestra Sec25 se desplaza en solitario en un brazo, se trata de la secuencia con más cambios, con un 21,7%, saliendo del mismo brazo que la

Sec31 ambas pancreáticas. La Sec82 y Sec83 permanecen unidas al igual que la Sec43 y Sec88, en diferentes zonas del árbol. Vemos la unión de muestras de páncreas que permanecen unidas, como la Sec37, Sec30, Sec28, Sec21, Sec22 con algo más de distancia y Sec20 con un desplazamiento mayor. Asimismo, observamos la unión en diferentes brazos de las demás secuencias de carcinoma de páncreas.

3. RESULTADOS.

3.1. Resultados del análisis de las mutaciones en las secuencias.

No podemos valorar en este trabajo la mutación en el gen KRAS en pacientes con carcinoma de páncreas porque las secuencias no alcanzaban a cubrir la zona de la mutación G12D/G12V (I. Roa et al 2013, S. Sho et al. 2017). Sin embargo, pacientes con carcinomas no pancreáticos si cubrían la zona en el fragmento 1 y el 4. En pacientes no pancreáticos, solo existe una secuencia que ha presentado una mutación en la posición 228 según la RefSeq, se trataba de la mutación GGT>GTT perteneciente a la Sec79 que es un paciente con un osteosarcoma.

Ningún paciente con carcinoma de vejiga ha presentado mutaciones a lo largo de la zona en estudio. En cambio, las muestras Sec43 y Sec88, acumulan un total de 29 cambios en su secuencia a lo largo de todo el fragmento. Estas secuencias pertenecen a carcinomas desconocidos. Difieren solo en que la Sec43 tiene la mutación G230A y la Sec88 la G620C, ambas con el 6.56% de mutaciones.

Destacan las mutaciones A376T encontrada en 7 secuencias, A262C en 17 secuencias, T261C en 10, A258G en 18 y 256D en 7, como las que más se repiten. En su mayoría pertenecen a pacientes con carcinoma de páncreas, aunque también afecta a otros tipos de cáncer como cerebro, pulmón y origen desconocido.

En los pacientes con carcinoma de páncreas, hubo cambios en las secuencias de repetición. Los cambios que más se han hallado son la delección en la posición 258, respecto a la secuencia RefSeq que se aprecia en 15 secuencias. La mutación de Adenina por Guanidina, en la misma posición, se aprecia en otras 4 secuencias. Igualmente, en la posición 261 se producen 9 cambios de Timina por Guanidina y una delección. En la siguiente posición, 262, se producen 17 cambios de Adenina por Citosina. Por último, en la posición 299 la Adenina es sustituida en dos secuencias por Citosina y en siete por Timina.

Las muestras pertenecientes a pacientes con carcinoma de páncreas destacan, la mayoría, por tener un alto porcentaje de mutaciones en el fragmento número 5, siendo el único fragmento en el que participan. De las 26 muestras secuencias, 19 contienen cambios, alcanzando un porcentaje de 21.7% en el caso de la Sec25, o un 18,3% en las Sec19 y Sec22. Por otro lado, 7 de las secuencias pancreáticas no contienen ninguna mutación.

Encontramos un total de 107 cambios en total en las 26 secuencias de pacientes con carcinoma de páncreas, con unas secuencias de 61.84pb de longitud media. En cambio, en los carcinomas no pancreáticos, se aprecian 84 cambios en 69 secuencias, teniendo una longitud media de 213.68pb. Con estos datos, podemos afirmar que padecen más mutaciones los pacientes con carcinoma de páncreas que lo pacientes con otros tipos de carcinomas.

3.2 Resultados del análisis descriptivo.

En el análisis descriptivo, se comparan las dos poblaciones, los pacientes con carcinomas pancreáticos y los pacientes con carcinomas no pancreáticos. El análisis se realiza de dos maneras, a través del número total de mutaciones y a través del tanto por cierto de mutaciones según la longitud de la secuencia. En ambos casos, las medias de las mutaciones que contienen las secuencias de los pacientes con carcinoma de páncreas son superiores a las mutaciones de los pacientes que sufren otro tipo de carcinomas.

Cuando hacemos el estudio descriptivo utilizando el número de mutaciones totales, la diferencia entre las medias de las dos poblaciones es reducida, siendo superior en los pacientes pancreáticos. Con una media de 5.63 ∓ 3.39 de desviación estándar en el grupo de pancreáticos frente a una media de 4.83 ∓ 8.89 de desviación estándar en el grupo de los no pancreáticos.

En cambio, si se utiliza el tanto por ciento teniendo en cuenta la longitud de las secuencias, la diferencia entre las medias es considerable. El grupo pancreático tiene una media de 9.39 ∓ 5.64 de desviación estándar, frente a una media de 1.63 ∓ 2.09 de desviación estándar en el grupo de los no pancreáticos.

Siguiendo con un test de ANOVA, o análisis de la varianza, obtenemos unos resultados que sugieren que en el estudio con el número total de secuencias, no hay diferencias estadísticamente significativas entre ambos grupos. Sin embargo, si el estudio se realiza a través del tanto por ciento de mutaciones teniendo en cuenta la longitud de las secuencias de ADN, ocurre todo lo contrario. Sí existen diferencias estadísticamente significativas entre ambos grupos, los pacientes con carcinomas pancreáticos y los no pancreáticos.

Estos resultados concuerdan con los boxplot de la figura 16 y 17. Se puede afirmar que a través del tanto por ciento de las mutaciones tienen en cuenta la longitud de la secuencia de ADN, que es superior en el grupo de pancreáticos. Los pacientes que tienen un carcinoma de páncreas acumulan más mutaciones genéticas en su secuencia de ADN que los pacientes que sufren otro tipo de carcinoma.

3.3 Resultados del análisis filogenético.

Al construir los árboles filogenéticos por ambos métodos, Neighbor-joining y máxima verosimilitud, observamos que los resultados son similares. Las secuencias Sec43 y Sec88, de origen desconocido, son las que más mutaciones acumulan y permanecen unidas en la misma rama en los cinco fragmentos del proyecto.

Las secuencias pancreáticas están representadas en los dos árboles del fragmento 5. Las secuencias que no contienen mutaciones que aparecen unidas en la misma rama. Mientras que las secuencias que contienen mutaciones aparecen dispersas por el árbol filogenético. Es decir, tienen comportamientos diferentes, permaneciendo cercanas a otras secuencias homólogas de cáncer de páncreas o acercándose a secuencias de otro tipo de carcinoma.

Las muestras Sec11, Sec12, Sec13, Sec62 y Sec93, cerebro, vejiga, pulmón y de origen desconocido se observan agrupadas en los árboles.

También podemos ver como se desplaza la Sec8 en el fragmento 2 y 3, secuencia que corresponde a un carcinoma desconocido. Lo mismo ocurre con la Sec79 de osteosarcoma.

En la Sec25 y Sec31, pertenecientes a pacientes con carcinoma de páncreas, observamos el mayor desplazamiento correspondiente con el mayor número de mutaciones, con respecto a las demás secuencias pancreáticas.

En las no pancreáticas se aprecia en los árboles de los cinco fragmentos la unión de las secuencias Sec82 y Sec83, que corresponden a pulmón. Incluso en el fragmento 2 y 3, se unen con las secuencias, Sec11, Sec12, Sec13 y Sec14 de cerebro y de origen desconocido.

También, las Sec43 y Sec88 de origen desconocido, con el mayor número de acumulación de mutaciones en su mayoría iguales, permanecen agrupadas en los cinco fragmentos.

Se observan resultados similares en ambos métodos de filogenia, Neighbor-joining y Máxima verosimilitud, aunque se aprecian mejor las distancias por el método de Máxima verosimilitud. Es decir, las muestras con los brazos más largos han tenido más cambios en su secuencia de ADN.

4. CONCLUSIÓN.

El bajo número de secuencias encontradas en los bancos de datos genómicos en la región del gen KRAS, es muy significativa. Ninguna secuencia perteneciente a carcinomas pancreáticos alcanzaba la zona de la mutación genética de interés G12D/G12V. Mutación con gran peso a la hora de padecer un carcinoma de páncreas.

El proyecto se ha llevado a cabo en los tiempos marcados en las PEC. Pero la escasez de secuencias de pacientes con carcinoma de páncreas ha llevado a ampliar el trabajo con pacientes que padecían diversos tipos de carcinomas, como se puede observar en el Anexo1. Esta situación ha beneficiado el proyecto, incrementando y enriqueciendo el trabajo de fin de máster. Pero a la vez, supone poco material para cualquier tipo de estudio sobre dicha mutación.

De igual manera, se ha puesto de manifiesto una falta de datos sobre las secuencias de los bancos de datos genómicos. Hubiera sido interesante haber tenido más información sobre los pacientes como la edad, hábitos, antecedentes familiares, alimentación, etc. Hubiera sido atrayente poder estudiar este tipo de datos junto con las mutaciones que presentan los pacientes en sus secuencias de ADN en el gen KRAS.

Ambos métodos de filogenia aportan resultados interesantes y agrupaciones similares, aunque con el método de máxima verosimilitud observamos mejor las distancias que guardan entre las secuencias genómicas y los cambios evolutivos de las mismas en el gen KRAS.

En resumen, se han hecho tres estudios dentro del trabajo de fin de master. Un análisis de las secuencias, un análisis descriptivo y un estudio poblacional a través de filogenia. En los tres estudios se ha podido comprobar el diferente comportamiento de las secuencias con mutaciones de los pacientes que padecían un carcinoma de páncreas.

Aunque los pacientes con carcinoma de páncreas tienen a acumular mutaciones de manera muy diversa, si bien solo participan en el fragmento 5, se confirma que tienen un mayor acúmulo de mutaciones en la región del gen KRAS. Situación que puede tener una gran relevancia.

4.1. Futuros estudios.

La mutación en el codón 12 es indicativa de una preneoplasia o cáncer en estadio temprano (D. Qing et al. 2004). Existen 10 años para una detección temprana, en una etapa potencialmente curativa y 5 años para la aparición de células con una alta difusión local y que pueden producir metástasis a distancia (C. Chan et al 2012). Con estas afirmaciones podemos sugerir que, serían necesarias técnicas de detección rápida del oncogén KRAS basadas en la detección de anticuerpos monoclonales EGFR (J. Hernández-Losa et al. 2011), donde el paciente puede tener la oportunidad de una mejora de su pronóstico. Ya existen en el mercado test de detección en heces y sangre donde, si se ampliara su uso, supondría un beneficio poblacional.

La carencia de secuencias en el gen KRAS ha enriquecido el trabajo de fin de máster, pero sería interesante para futuros estudios aumentar el número de secuencias en los servidores genómicos dado que supone poco material para cualquier tipo de estudio sobre dicha mutación. Este incremento de secuencias en el gen KRAS favorecería y mejoraría posteriores estudios.

Son muy interesantes cambios reincidentes encontrados en los fragmentos de las secuencias del gen KRAS de los pacientes con carcinoma de páncreas. Estos cambios se localizan en la posición 256, 258, 261, 262 y 376 según la secuencia RefSeq, cambios que podrían ser de gran interés para futuros estudios. Quizá para estudios de nuevas dianas terapéuticas que mejoren el diagnóstico precoz y la creación de nuevos fármacos que mejoren la supervivencia de los pacientes.

5. GLOSARIO.

ADN: Ácido desoxirribonucleico, constituye el material genético de los seres vivos.

EGFR: Receptor del factor de crecimiento epidérmico.

KRAS: Nombre de la proteína que tiene como función controlar el crecimiento, maduración y muerte celular.

Mut: Mutación genética.

NCBI: Navegador genómico del Centro Nacional de Información Biotecnológica.

Pb: Par de base, aminoácido en la secuencia de ADN.

RefSeq: Secuencia consenso de la entrada de múltiples secuencias en la misma región en los bancos de datos genómicos.

Sec: Secuencia de ADN.

UCSC: Navegador genómico de la Universidad de California en Santa Cruz

6. BIBLIOGRAFÍA.

1. A. Utrillas Martínez, M. López Bañeres y J. M. del Val Gil. Cáncer de páncreas. Bol oncol. del área sanitaria de Teruel. ISSN 1579-6094.
2. Alyssa M Krasinskas, A James Moser, Burcu Saka, N Volkan Adsay, and Simion I Chiosea. KRAS mutant allele-specific imbalance is associated with worse prognosis in pancreatic cancer and progression to undifferentiated carcinoma of the pancreas. *Mod Pathol.* 2013 October ; 26(10): 1346–1354.
3. C. Chan, O. Santes. Neoplasias pancreáticas: Avances. *Revista de Gastroenterología de México* 2012;77(Supl 1):108-111.
4. C. Hu, N. Hart, C. Polley, PhD, Gnanaolivu, Shimelis, PhD, Lee, Lilyquist, Jie Na, Raymond Moore, BS, S. O. Antwi, PhD. Association Between Inherited Germline Mutations in Cancer Predisposition Genes and Risk of Pancreatic Cancer. *JAMA.* 2018 June 19; 319(23): 2401–2409.
5. Campbell, P.J.; Yachida, S.; Mudie, L.J.; Stephens, P.J.; Pleasance, E.D.; Stebbings, L.A.; Morsberger, L.A.; Latimer, C.; McLaren, S.; Lin, M.L.; et al. The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* 2010, 467, 1109–1113.
6. Carlos Augusto Meneses Escobar, Lizeth Vanessa Rozo Murillo, Jhenifer Franco Soto. Tecnologías bioinformáticas para el análisis de secuencias de ADN. *Scientia et Technica* Año XVI, No 49, diciembre de 2011. Universidad Tecnológica de Pereira. ISSN 0122-1701.
7. Carlos Peña. Métodos de inferencia filogenética. *Rev. Peru. Biol.* 18(2): 265-267 (Agosto 2011).

8. Daniela Dias Xavier. Estrategia de solución al problema de la anotación de secuencias de ADN mediante metodología CommonKADS. Facultad de informática Universidad Complutense de Madrid.
9. De-Qing Mu, You-Shu Peng, Qiao-Jian Xu. Values of mutations on K-ras oncogene at codon 12 in detection of pancreatic cancer: 15-years experience. *World J Gastroenterol.* 2004; 10(4): 471-475.
10. Donald D. Rao, Xiuquan Luo, Zhaohui Wang, Christopher M. Jay, Francis C. Brunicardi, William Maltese, Luisa Manning, Neil Senzer, John Nemunaitis. KRAS mutant allele-specific expression knockdown in pancreatic cancer model with systemically delivered bi-shRNA KRAS lipoplex. *Plos One* 2018.
11. E. Takai, Y. Totoki, H. Nakamura, C. Morizane, S. Nara, N. Hama, M. Suzuki, E. Furukawa, M. Kato, H. Hayashi, T. Kohno, H. Ueno, K. Shimada, T. Okusaka, H. Nakagama, T. Shibata, S. Yachida. Clinical utility of circulating tumor DNA for molecular assessment in pancreatic cancer. *Nature* 2016.
12. Elvira Pelosi, Germana Castelli and Ugo Testa. Pancreatic Cancer: Molecular Characterization, Clonal Evolution and Cancer Stem Cells. *Biomedicines* 2017, 5, 65.
13. G. Argilés, R. Dienstmann, M. E. Élez y J. Tabernero. Principales alteraciones genéticas en el cáncer colo-rectal. Servicio de Oncología Médica Hospital Universitario Vall d'Hebrón, Barcelona (Agosto,2012).
14. Iván Roa, Tamara Sánchez, Alejandro Majlis, Kurt Schalper. Mutación del gen KRAS en el cáncer de colon y recto. *Rev. méd. Chile* vol.141 no.9 Santiago set. 2013.
15. James R. Eshleman, Alexis L. Norris, Yoshihiko Sadakari, Marija Debeljak, Michael Borges, Colleen Harrington, Elaine Lin, Aaron Brant, Thomas Barkley, J. Alejandro Almario, Mark Topazian, James Farrell, Sapna Syngal, Jeffrey H. Lee, Jun Yu, Ralph H. Hruban, Mitsuro Kanda, Marcia Irene Canto, and Michael Goggins. KRAS and GNAS

Mutations in Pancreatic Juice Collected From the Duodenum of Patients at High Risk for Neoplasia Undergoing Endoscopic Ultrasound. *Clin Gastroenterol Hepatol*. 2015 May; 13(5): 963–969.

16. Javier Hernández-Losa, Julián Sanz, Stefania Landolfi, Fernando López-Ríos, José Palacios, María Dolores Bautista, Eduardo Díaz-Rubio, Josep Taberner, Jesús García Foncillas y Santiago Ramón y Cajal. Recomendaciones para la determinación de mutaciones de K-RAS en cáncer de colon. *Rev Esp Patol*. 2012;45(2):76---85.

17. K.A. Palacio-Rúa, L.F. Isaza-Jiménez, E. Ahumada-Rodríguez, H. Ceballos-García y C.M. Muñetón-Peña. Análisis genético en APC, KRAS y TP53 en pacientes con cáncer de estómago y colon. *Revista de Gastroenterología de México*. 2014;7977(Supl 1):108-111.

18. Manuel Hidalgo, M.D. Pancreatic Cancer. *Medical Progress. n engl j med* 362;17 nejm.org april 29, 2010.

19. O. Sánchez Becerro. Algoritmos de programación dinámica con R para resolver problemas de alineamiento de secuencias. TFM 2014. Universidad Complutense de Madrid.

20. Patricia Alejandra López Penza, Luis Ruso Martínez. Cáncer de páncreas. Epidemiología de su mal pronóstico. *Rev Méd Urug* 2017; 33(3):122-128.

21. Qing Wanga,b, Raghothama Chaerkadyc,d, Jian Wua,b, Hee Jung Hwanga,b, Nick Papadopouloua,b, Levy Kopeloviche, Anirban Maitraf,g, Hanno Matthaeif,g, James R. Eshlemanf,g, Ralph H. Hrubanf,g, Kenneth W. Kinzlera,b, Akhilesh Pandeyc,f,g,h,1, and Bert VogelsteinaMutant proteins as cancer-specific biomarkers *PNAS*. February 8, 2011. vol. 108 no. 6.

22. Shonan Sho, Colin M. Court, Stephen Kim, David R. Braxton, Shuang Hou,V. Raman Muthusamy, Rabindra R. Watson,Alireza Sedarat, Hsian-Rong Tseng, and James S.

Tomlinson. Digital PCR Improves Mutation Analysis in Pancreas Fine Needle Aspiration Biopsy Specimens. Published online 2017 Jan 26.

23. Yachida, S. Jones, S. Bozic, I. Antal, T. Leary, R. Fu, B, Kamiyama, M. Hruban, R.H. Eshleman, J.R. Nowak, M.A. et al. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* 2010, 467, 1114–1117.

7. ANEXOS.

Anexo1. Tabla descriptiva de las secuencias de ADN con la longitud de las secuencias, mutaciones en cada fragmento y tipos de cáncer.

Sec	Mut	pbGenBank	pb	Mut%	Tipo_Cancer	Frag1	Frag2	Frag3	Frag4	Frag5
Sec1	0	5765	442	0	RefSeq					
Sec2	0	581	442	0	Unknown					
Sec3	0	178006	287	0	Unknown					
Sec4	0	125756	340	0	Unknown					
Sec5	0	648	108	0	Brain					
Sec6	0	567	442	0	Unknown					
Sec7	1	390	108	0,93	Colon and lung					A294G
Sec8	1	4418	287	0,35	Unknown		A376C			
Sec9	0	1418	442	0	Testis					
Sec10	0	1333	442	0	Lung					
Sec11	3	1693	442	0,68	Brain	204D 205D	A376T			
Sec12	1	567	442	0,23	Unknown		A376T			
Sec13	1	607	442	0,23	Unknown		A376T			
Sec14	1	607	442	0,23	Unknown		A376T			
Sec15	1	161	60	1,67	Pancreatic					258D
Sec16	4	162	60	6,67	Pancreatic					258D T261C A262C A294T
Sec17	4	163	60	6,67	Pancreatic					258D A262C G284C A299T
Sec18	3	166	60	5	Pancreatic					258D T261C A262C A294T
Sec19	11	171	60	18,3	Pancreatic					A251G G253A A256T 258D T261C A262C G284C A296G A299C 300I C302T
Sec20	9	160	60	15	Pancreatic					249D A251G 258D A262C A263T A267G A273T A298C A299T
Sec21	7	163	60	11,7	Pancreatic					256D A258G T261C A262C A267G A296G A299T

Sec22	11	156	60	18,3	Pancreatic					A256C C257G 258D A262C A267G A295G A296C A298C A299T 300D 301G
Sec23	0	131	60	0	Pancreatic					
Sec24	3	131	60	5	Pancreatic					258D A262C A299T
Sec25	13	184	60	21,7	Pancreatic					G245C T248G A251G G256C C257G 258D C260A T261C A262C A263T T264G A270G G284C A296G A298C A299C G302C
Sec26	0	168	60	0	Pancreatic					
Sec27	7	162	60	11,7	Pancreatic					258D T261C A262C T264G G284C A296G A301G
Sec28	5	157	60	8,33	Pancreatic					246D A258G A262C A267G A299T
Sec29	4	165	60	6,67	Pancreatic					A258G 261D A262C A299T
Sec30	4	134	60	6,67	Pancreatic					256D A258G A262C A267G
Sec31	8	173	60	13,3	Pancreatic					G253A A256C C257G 258D T260C A262T G284C 300I
Sec32	2	153	60	3,33	Pancreatic					258D A273T
Sec33	5	163	60	8,33	Pancreatic					256D A258G T261C A262C T264G
Sec34	0	130	60	0	Pancreatic					
Sec35	0	163	60	0	Pancreatic					
Sec36	0	132	60	0	Pancreatic					
Sec37	3	130	60	5	Pancreatic					258D A262C A299T
Sec38	0	131	60	0	Pancreatic					
Sec39	3	127	60	5	Pancreatic					258D T261C A262C

Sec40	0	126	108	0	Pancreatic					
Sec41	1	123	108	0,93	Gallbladder					A284C
Sec42	0	49663	108	0	Sarcoma					
Sec43	29	865	442	6,56	Unknown	T220C G230A G236A G241A C246T G284C	A361T, G368A, G379A, G387A, 420D 416I	A305G C453T C472G C476A 396D	T213C G485A G500T A520G A575C C578A A598G T604G T613C 570D 571D 625D 626D	
Sec44	0	154	60	0	Colorectal					
Sec45	0	891	442	0	Protate					
Sec46	0	186	108	0	Colorectal					
Sec47	0	53713	0	0	Unknown					
Sec48	0	136	108	0	Lung					
Sec49	1	136	108	0,93	Lung	G227T				
Sec50	0	159	0	0	Colorectal					
Sec51	0	169	0	0	Colorectal					
Sec52	0	213	0	0	Colorectal					
Sec53	0	210	108	0	Cervical					
Sec54	0	223	0	0	Cervical					
Sec55	0	133	101	0	Bladder					
Sec56	0	135	101	0	Bladder					
Sec57	0	135	101	0	Bladder					
Sec58	0	134	101	0	Bladder					
Sec59	0	135	101	0	Bladder					
Sec60	0	133	101	0	Bladder					
Sec61	0	139	101	0	Bladder					
Sec62	0	132	101	0	Bladder					
Sec63	0	135	101	0	Bladder					
Sec64	0	139	101	0	Bladder					
Sec65	0	135	101	0	Bladder					
Sec66	0	134	101	0	Bladder					
Sec67	0	132	101	0	Bladder					
Sec68	0	133	101	0	Bladder					
Sec69	0	133	101	0	Bladder					
Sec70	0	135	101	0	Bladder					
Sec71	0	134	101	0	Bladder					
Sec72	0	132	101	0	Bladder					
Sec73	0	133	101	0	Bladder					
Sec74	0	135	101	0	Bladder					
Sec75	0	133	101	0	Bladder					
Sec76	0	135	101	0	Bladder					
Sec77	0	135	101	0	Bladder					
Sec78	0	139	101	0	Bladder					
Sec79	2	113	108	1,85	Osteosarcoma	G228T G214C				

Sec80	1	113	108	0,93	Lung	G227T				
Sec81	0	111	60	0	Lung					
Sec82	5	450	442	1,13	Lung	C266G G268C	A376T		G620C	
Sec83	5	450	442	1,13	Lung	G227T C266G G268C			G620C	
Sec85	1	5775	442	0,23	Unknown	G227T				
Sec86	0	258	178	0	Cervical					
Sec87	0	35590	0	0	Unknown					
Sec88	29	5458	442	6,56	Unknown	T220C, G236A, G241A, C246T G284C,	A361T G368A G379A G387A 396D 420D 416I	A305G C453T C472G C476A 396D	T213C G485A G500T A520G A575C C578A A598G T604G T613C 570D 571D 625D 626D	
Sec89	0	52675	0	0	Unknown					
Sec90	0	5889	442	0	Unknown					
Sec91	0	27634757	0	0	Unknown					
Sec92	1	111	108	0,93	Myelogenous leukemia	A286T				
Sec93	1	560	178	0,56	Lung		A376T			
Sec94	0	560	178	0	Lung					
Sec95	0	912	442	0	Unknown					
Sec96	0	5299	442	0	Unknown					

Sec	A296G	A298C	A299C	A299T	300D	300I	A301G	C302T	A305G	A361T	G368A	A376C	A376T	G379A	G387A	396D	420D	416I	C453T	C472G	C476A	G485A	G500T	A520G	570D	571D	A575C	C578A	A598G	T604G	T613C	G620C	625D	n° mut	Sec	
Sec7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	Sec7
Sec8	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	Sec8
Sec11	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	Sec11	
Sec12	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	Sec12	
Sec13	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	Sec13	
Sec14	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	Sec14	
Sec15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	Sec15	
Sec16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	Sec16		
Sec17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	Sec17		
Sec18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	Sec18		
Sec19	1	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	Sec19	
Sec20	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	Sec20		
Sec21	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	Sec21		
Sec22	1	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	Sec22		
Sec24	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	Sec24		
Sec25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	Sec25		
Sec27	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	Sec27		
Sec28	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	Sec28		
Sec29	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	Sec29		
Sec30	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	Sec30		
Sec31	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	Sec31		
Sec32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	Sec32		
Sec33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	Sec33		
Sec37	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	Sec37		
Sec39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	Sec39		
Sec41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	Sec41		
Sec43	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	29	Sec43		
Sec49	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	Sec49	
Sec79	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	Sec79		
Sec80	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	Sec80		
Sec82	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	5	Sec82		
Sec83	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	5	Sec83	
Sec85	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Sec85
Sec88	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	29	Sec88		
Sec92	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Sec92
Sec93	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Sec93
N° muta	4	2	2	7	2	1	2	1	2	2	2	6	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3	2	190	N° mut
	A296G	A298C	A299C	A299T	300D	300I	A301G	C302T	A305G	A361T	G368A	A376C	A376T	G379A	G387A	396D	420D	416I	C453T	C472G	C476A	G485A	G500T	A520G	570D	571D	A575C	C578A	A598G	T604G	T613C	G620C	625D	n° mut		