

BT1013 Análisis de Biología Computacional
ENTREGABLE DE LA SITUACIÓN PROBLEMA

Integrantes

Carlos Manuel Brito Diaz | A01283561
Alan Eduardo Barrón Cadena | A00829058
Luis Daniel Cano Irigoyen | A00827178
Gerardo Andrés Castañon | A01283912

Profesores

Dr. Jesús Hernández Pérez
Dr. Luis Miguel Figueroa Galván
Dr. Edgar Acuña Gonzalez

SITUACIÓN PROBLEMA:

Identificar las diferencias genéticas de pacientes con cáncer de colon de distintas edades (pacientes jóvenes [< 35 años] y adultos de edad avanzada [> 50 años]) usando bases de datos y herramientas de biología computacional.

Introducción

En 2018, un estudio hecho por la Sociedad Americana Contra el Cáncer encontró que personas nacidas en 1990, que hubieran tenido 28 años en 2018, tenían el doble de riesgo de cáncer de colon, y cuádruple de cáncer de recto. El estudio encontró un aumento anual de 2.4% en la incidencia de casos de cáncer de colon en adultos entre los 20 y los 39 años, mientras que en adultos arriba de 55 años, las tasas de incidencia estaba disminuyendo.

De forma general, el cáncer es el nombre que se le da a todo un conjunto de enfermedades relacionadas. Hasta la fecha, se han identificado más de 100 tipos distintos de cáncer. Sin embargo, todos los tipos de cáncer presentan una división celular descontrolada con el potencial de propagarse o invadir el resto del cuerpo. Normalmente, cuando las células en el cuerpo se vuelven muy viejas o se dañan, son eliminadas y reemplazadas por nuevas células. Cuando el cáncer se desarrolla, se observa una interrupción en el proceso normal de división celular, lo que hace que las células viejas no sean eliminadas, y que se forman células nuevas cuando no son necesarias. Estas células extras se dividen sin parar y pueden llegar a formar crecimientos llamados tumores que tienen el potencial de invadir el resto del cuerpo, en contraste con los tumores benignos.

Se estima que cada año, en el mundo, hay más de 7 millones de defunciones atribuidas al cáncer. Esto significa que la probabilidad de morir por cáncer es de 12% (1 en cada 8). Tan solo en México, el cáncer constituye un 12% de las defunciones totales, siendo la 3era causa de muerte para la población general. Uno de los problemas fundamentales de la lucha contra el cáncer, y la razón detrás de los aumentos en las incidencias del cáncer en la población es la eliminación de otras causas de muerte. Hace menos de un siglo, la expectativa de vida era alrededor de 50 años. Hoy en día es casi 79, y la edad promedio de muerte por cáncer es de 72 años. Como lo pone Alberts B, Johnson:

If a human could live long enough, it is inevitable that at least one of his or her cells would eventually accumulate a set of mutations sufficient for cancer to develop.

Si bien la conclusión es que el cáncer eventualmente se presentará en personas lo suficientemente viejas, existe una preocupación en el aumento de las incidencias de distintos tipos de cáncer en personas jóvenes (<35 años). Uno de estos tipos de cáncer es el cáncer de colon. El cáncer de colon es un tipo de cáncer colorrectal que comienza en el intestino largo. Este tipo de cáncer suele comenzar como pequeños grupos de células llamadas pólipos que se forman en el interior del colon. Algunos tipos de pólipos pueden llegar a convertirse en cáncer con el tiempo, pero el cambio depende del tipo de pólipo. Los dos tipos de pólipos principales son:

- **Pólipo adenomatoso:** Estos pólipos pueden convertirse en cáncer. Casi un 40% de las veces que se encuentra un pólipo pre-canceroso, es un adenoma.
- **Polipo hiperplásico:** Este tipo de pólipos son más comunes, pero generalmente no son pre-cancerosos.

Si el cáncer se forma en un pólipo, puede crecer dentro de las paredes del colon con el paso del tiempo. El cáncer de colon comienza en la mucosa del intestino, y puede crecer hacia afuera hasta haberse propagado hacia todas las capas. Cuando las células cancerígenas se encuentran en la pared intestinal, pueden propagarse a vasos sanguíneos o linfáticos. De ahí, el cáncer puede viajar a partes distantes del cuerpo. La etapa en la que se encuentra el cáncer de colon depende de qué tan profundamente se haya desarrollado en los muros intestinales, y si se ha esparcido a otras partes del cuerpo.

Si bien los pólipos pueden llegar a convertirse en cáncer, el periodo de tiempo en el que esto sucede es alrededor de 10 a 15 años. Hoy en día existen muchos esfuerzos por partes de distintas organizaciones para el diagnóstico y tratamiento en todas las etapas del cáncer de colon. Este tipo de cáncer puede, usualmente, ser detectado en etapas tempranas con pruebas de detección

regulares. Es en estas etapas iniciales cuando es más tratable y los médicos pueden extirpar los pólipos antes de que puedan convertirse en cáncer. Algunos de los métodos utilizados para diagnosticar el cáncer de colon actualmente son:

- **Pruebas de sangre oculta en heces:** Esta prueba es llevada a cabo para detectar rastros de sangre en las heces de una persona. La presencia de sangre indica que puede que haya algún tipo de sangrado en el tubo digestivo. Mientras que el sangrado puede tener una cierta variedad de causas como úlceras o hemorroides, si la persona presenta otros síntomas explicados por cáncer colorrectal, los resultados son más conclusivos.
- **Examen de sangre:** Esta prueba usualmente busca medir distintos elementos de la sangre. Por ejemplo, una cuenta de células que revele un número bajo de células rojas en la sangre indica anemia. Algunas personas con cáncer colorrectal se vuelven anémicas ya que el tumor ha estado sangrando por mucho tiempo.
- **Colonoscopia diagnóstica:** Esto usualmente se lleva a cabo cuando se obtienen resultados anormales en el diagnóstico inicial. Para este examen, el doctor inspecciona toda la longitud del colon y el recto con un colonoscopio. Instrumentos especiales pueden ser pasados por el colonoscopio para remover cualquier área que parezca estar desarrollando pólipos, o para tomar muestras del tejido.
- **Biopsia:** En la biopsia, el doctor remueve piezas pequeñas de tejido para analizarlo. En muchas ocasiones, los doctores buscan cambios genéticos específicos en las células con cáncer que puedan afectar como se trata. Por ejemplo, los doctores típicamente buscan cambios en los genes KRAS, NRAS y BRAF.

Existen muchas otras formas distintas de diagnosticar el cáncer de colon, como exámenes de resonancia magnética (MRI), ultrasonidos, tomografías por emisión de positrones (PET), e incluso formas de rastrear el progreso del cáncer por el cuerpo, como rayos-x para determinar si el cáncer se ha esparcido a los pulmones. Sin embargo, los investigadores siempre están buscando nuevas tecnologías que puedan ser utilizadas para mejorar su diagnóstico y entendimiento de las causas del cáncer. En los últimos años, por ejemplo, se ha visto un incremento en el interés en las Inteligencias Artificiales y el Aprendizaje Automático para la detección de pólipos en pacientes y tendencias cancerígenas en muestras de tejido de pacientes que presenten síntomas. En 2019, un equipo de investigadores utilizó la inteligencia artificial para analizar más de 26,000 fotogramas de tejido colorrectal y lograron diferenciar tejidos con y sin cáncer con 100% de precisión en el estudio piloto.

Uno de los aspectos más importantes para el estudio del cáncer en general, es el análisis de la información disponible. Comenzando con el Proyecto Genoma Humano fundado en los 90s, en las últimas dos décadas hemos presenciado una explosión en la creación y uso de herramientas de secuenciado de ADN. Estas herramientas han permitido que científicos puedan recolectar cantidades enormes de información sobre los genomas de tejidos y células normales y cancerígenas, y tratar de entender como es lo que causa el cáncer e intentar crear modelos para predecir cómo es que este progresa.

El poder de las computadoras para trabajar con cantidades muy grandes de información como datasets facilita que investigadores puedan hacer muchos tipos distintos de análisis del comportamiento y expresión de genes a partir de datasets que pueden llegar a abarcar décadas. Debido a que, en muchos aspectos, el análisis genómico es similar al trabajo que las personas que hacen data science llevan a cabo, muchas de las herramientas que se utilizan son las mismas. Algunas de las herramientas más utilizadas son Python, R, Bioconductor y Galaxy. Este tipo de herramientas permiten trabajar con las bases de datos que son provistas por una variedad de repositorios como NCBI, DDBI o EMBL.

En la lucha contra el cáncer, específicamente, se han llevado a cabo estudios a lo largo de las décadas con el objetivo de entender los factores de riesgo del cáncer y entender cómo es que la expresión genética de distintos tipos de cáncer en distintos tipos de tejidos nos puede ayudar a entender cómo combatirlo y como diseñar mejores modelos de prevención y diagnóstico. Sin embargo, uno de los usos de la inteligencia artificial en esta área se ha visto a través de la construcción de modelos predictivos que puedan predecir vulnerabilidades en el cáncer a partir de los perfiles genéticos de este.

Si bien este tipo de análisis requieren de una cantidad más grande de recursos y de implementaciones muy complejas de las herramientas disponibles, solo ver las diferencias que distintos tipos de tejidos con y sin cáncer tienen nos puede decir mucho sobre cómo es que este se desarrolla y entender si existe alguna predisposición a distintos tipos de cáncer bajo condiciones de edad distintas, como se hará en este análisis.

Metodología

Para poder llevar a cabo un análisis genético de la expresión del cáncer en personas de distintas edades en una herramienta como R, debemos de lograr conseguir alguna base de datos que tenga los registros que se buscan. Si se contaran con más recursos para llevar a cabo un estudio de este tipo, se podría aportar a dicha base de datos al añadir nuestras propias mediciones

y datos. Sin embargo, estaremos utilizando bases de datos públicas compiladas por instituciones externas.

En nuestro caso, estaremos trabajando con la base de datos GSE25071, compilada por el Hospital de la Universidad de Oslo. Originalmente, la base de datos fue compilada con el propósito de investigar genes asociados con el aumento del riesgo del cáncer colorrectal en personas jóvenes. Es importante mencionar que se excluyeron pacientes con síndromes de predisposición hereditaria al cáncer. Esta base de datos contiene las expresiones genéticas de muestras de cáncer colorrectal de pacientes jóvenes y ancianos. En nuestra investigación, encontramos algunos artículos que habían utilizado la base de datos. Uno de los más interesantes para nuestra investigación fue el de *Identification of key genes and pathways involved in microsatellite instability in colorectal cancer*, escrito por Chaoran Yu, Hiju Hong, Sen Zhang, Yaping Zong, Junjun Ma, Aiguo Lu, Jing Sun, y Minhua Zheng, con el PMID 30664178. El objetivo de este artículo es identificar genes y vías de señalización asociadas con la inestabilidad microsatelital, cuya presencia se ha denominado como importante para los casos de cáncer colorrectal.

Si bien la base de datos contiene la información necesaria, se tuvieron que hacer algunas modificaciones para poder analizarla de la forma correcta. La base de datos consiste en dos tablas: una con 10,548 genes en las filas asociados a 50 distintas muestras en las columnas. La segunda es una tabla con las 50 muestras en las filas, y las columnas con los parámetros de edad, género, etapa, ubicación (si es en el colon o en el recto), la estabilidad satelital, y el tejido de la muestra.

Ahora, la primera modificación que llevamos a cabo fue dividir las muestras entre las que pertenecen a jóvenes menores a 35 años, y las de adultos mayores a los 50 años. Una vez que se tengan ambos índices definidos, se crean dos tablas nuevas: una con las muestras de jóvenes en las columnas y todos los genes en las filas, y otra con las muestras de ancianos en las columnas y los genes en las filas.

Con estas dos tablas de datos de jóvenes y adultos, debemos entonces tener una forma sobre la cual hacer nuestro análisis y tener visibilidad sobre datos estadísticos pertinentes a los datos. Para esto, decidimos crear una nueva tabla con 4 columnas: los estimados de las medias de los jóvenes en cada gen, los estimados de las medias de los ancianos en cada gen, los p-valores obtenidos, y al final la diferencia de medias entre ambos grupos de edad. En este caso, el objetivo de analizar el p-value para cada caso es tener una forma de ver la expresión de cada gen en cada muestra dependiendo de la edad. Entre menor sea el p-value, mayor es la certeza de que ese gen es significativo y de que tiene una relevancia para el comportamiento del cáncer fuera de

cualquier error estadístico. Por otro lado, la diferencia de medias nos da información sobre el comportamiento de la expresión de los datos. Las diferencias de medias que hayan resultado ser mayor a cero indican genes que resultaron ser mayormente expresados aun después de haber restado las medias de los datos de pacientes viejos contra las medias de pacientes jóvenes.

Finalmente, filtramos aún más los datos para solo quedarnos con los p-values que fueran menores a 0.05. Esto nos da un buen margen de error, con solo u 5% de probabilidad de que los resultados de la significancia se deban a algún elemento de azar. Después, ordenamos los datos para mostrarnos las filas cuyos p-values fueran mucho menores al comienzo, es decir, de forma descendente. Adicionalmente, dividimos estos datos entre los que tenían diferencias de medias positivas (jóvenes) y diferencias negativas (adultos), para poder ver cuáles genes son los mayormente expresados en cada grupo.

Resultados

Con la metodología anterior en mente, en este apartado se encuentran los resultados de nuestra investigación, en los cuales se encuentran primeramente la lista de los 75 genes más afectados (genes diferencialmente expresados) de la población juvenil (Tabla 2.1), seguido de la lista de los 75 genes más afectados (genes diferencialmente expresados) de la población adulta (Tabla 2.2)

JÓVENES
FIG4
EIF1B
SF1
BRPF1
HPF1
CEP350
KHDRBS1
ECHDC3
CNNM3
SAMD1
RBFOX2
C6orf136
CAMK1D
AXIN2
PIK3R4
PRPF19

**TM2D3
PRKCSH
EIF3A
PCDHGA1
ZNF222
ZNF766
CCND1
SCRN1
BCL7A
MTPN
FAM109B
MED29
NKAP
UBL7
VGLL4
SH3BP5L
RRAGA
BAX
KAT6B
CSNK2B
ARSD
TMEM99
PTPN23
TRMT2A
ZFAND6
NKIRAS2
ZNF562
LSM6
FAM53B
TSPO
GOSR1
PAK6
NR1H2
FAF1
UBTF
PURA
WDR18
IQGAP2
CNPY3
TRIM62
BBX
MLF2
KDM4A
KLHL8**

SAMM50
HNRNPUL1
ALKBH5
KDM5C
JAK1
CST3
DHPS
TBL1XR1
LDB1
EIF2AK2
DBNDD1
RECQL
RAB11B
ZNF398
DCBLD1

Tabla 2.1: Lista de los 75 genes más afectados de la población juvenil

ADULTOS
UBAP1 TMX2 GCH1 EIF1AY FUT10 HPCA LMF2 DRD4 CCNL2 MTRF1 GEM RPS4Y1 MFSD14B CRYGA MKKS BCL2L2 CRISPLD2 IDH3B CSRP1 VASP MCL1

**OR8A1
DDX3Y
C6orf62
COASY
SLC20A1
VCAM1
OGT
FNDC4
CALB2
EYA2
N4BP2L2
NME7
AAAS
KRTAP20-2
METTL17
C2
MMP19
RNF130
ATG4B
MPC1
CORO1C
NFIL3
SMIM7
MEIS1
ATP5O
IL3RA
TUFM
SEMA3G
SAMD12
FXD1
CHRNA2
SCRT1
RAD51B
SNRNP40
GMCL1
DUSP5
BIN3
NME3
GLRX5
ITGAX
STEAP1
PRNP
GPR84
STEAP1B**

TBRG1
RAB26
NID2
ZFP69B
KCTD10
PRKAG1
AKT1
SLC22A14
PRADC1
ALDH1B1

Tabla 2.2: Lista de los 75 genes más afectados de la población adulta

Estas tablas se ordenaron después utilizando

También podemos observar las rutas con las que se encuentran relacionadas los genes más afectados de la población joven (Figura 2.3) y los genes más afectados de la población adulta (Figura 2.4)

Gene Set Name [# Genes (K)]	Description	# Genes in Overlap (k)	k/K	p-value ?	FDR q-value ?
GSE24210_CTRL_VS_IL35_TREATED_TCONV_CD_CD4_TCELL_UP [200]	Genes up-regulated in T conv cells: control versus treated with IL35.	6	<div></div>	2.76 e-6	1.98 e-2
REACTOME_CHROMATIN_MODIFYING_ENZYMES [274]	Chromatin modifying enzymes	6	<div></div>	1.66 e-5	2.79 e-2
GSE34156_UNTREATED_VS_6H_TLR1_TLR2_LIG LIGAND_TREATED_MONOCYTE_DN [179]	Genes down-regulated in monocytes (6h): untreated versus M. tuberculosis 19 kDa lipopeptide.	5	<div></div>	2.76 e-5	2.79 e-2
GSE7509_UNSTIM_VS_FCGRIIB_STIM_DC_DN [182]	Genes down-regulated in dendritic cells: untreated versus anti-FcgRIIB.	5	<div></div>	2.98 e-5	2.79 e-2
GSE20484_MCSG_VS_CXCL4_MONOCYTE_DERIVE IVED_MACROPHAGE_DN [195]	Genes down-regulated in monocyte derived macrophages: CSF1 [GeneID=1435] versus PF4 [GeneID=5196].	5	<div></div>	4.15 e-5	2.79 e-2

Figura 2.3 Rutas relacionadas con los 75 genes más afectados de la población joven






Gene Set Name [# Genes (K)]	Description	# Genes in Overlap (k)	k/K	p-value ?	FDR q-value ?
GSE34515_CD16_NEG_VS_POS_MONOCYTE_UP [200]	Genes up-regulated in monocytes: CD16- versus CD16+.	7		1.38 e-7	9.84 e-4
GSE19923_E2A_KO_VS_E2A_AND_HEB_KO_DP_TP_THYMOCYTE_DN [200]	Genes down-regulated in double positive thymocyte: TCF3 [GeneID=6929] knockout versus TCF3 and TCF12 [GeneID=6929;6938] knockout.	6		2.76 e-6	6.59 e-3
GSE25890_CTRL_VS_IL33_IL7_TREATED_NUOCUOCYTES_DN [200]	Genes down-regulated in nuocytes: control versus treated with IL7 and IL33 [GeneID=3574;90865].	6		2.76 e-6	6.59 e-3
GSE25677_MPL_VS_MPL_AND_R848_STIM_BCELL_CELL_UP [164]	Genes up-regulated in B lymphocytes immunized with: monophosphoryl lipid A (MPL) versus MPL and imiquimod [PubChem=13982876].	5		1.81 e-5	8.37 e-3
GSE21927_SPLEEN_VS_4T1_TUMOR_MONOCYTE_TE_BALBC_UP [190]	Genes up-regulated in CD11b+ cells from BALB/c mice bearing 4T1 mammary carcinoma: spleen of BALB/c mice: spleen versus tumor infiltrating monocytes.	5		3.67 e-5	8.37 e-3

Figura 2.4 Rutas relacionadas con los 75 genes más afectados de la población adulta

Discusión

Dos de los criterios que hemos utilizado para las comparaciones y evaluaciones de los genes en nuestro análisis ha sido el p-value y la diferencia de medias. Es importante tomar ambos en cuenta ya que mientras que el p-value nos dice la confiabilidad de la comparación entre nuestras muestras, la diferencia de medias nos da información de la actividad celular de uno de los grupos con respecto al otro. Es decir, si la diferencia de medias resulta negativa, esto significa que hay una mayor actividad de un cierto gen en el grupo de pacientes jóvenes, mientras que una diferencia de medias positiva implica que el gen tuvo una mayor actividad en pacientes viejos.

Con esto en mente, realizamos el análisis de uno de los genes más expresados en la población joven; el gen JAK1, como se muestra en la Figura 2.5



Figura 3.1 Vías de señalización del gen JAK1 con la base de datos MSigDB

Los parámetros de bases de datos contra las que se hizo la comparación fueron las siguientes:

H : hallmark gene sets

CP : Canonical pathways

CP :BIOCARTA: BioCarta gene sets

CP :KEGG: KEGG gene sets

CP :PID: PID gene sets

CP :REACTOME: Reactome gene sets

C6 : oncogenic signatures

C7 : immunologic signatures

Visto en Cosmic 3D (Figura 2.6):

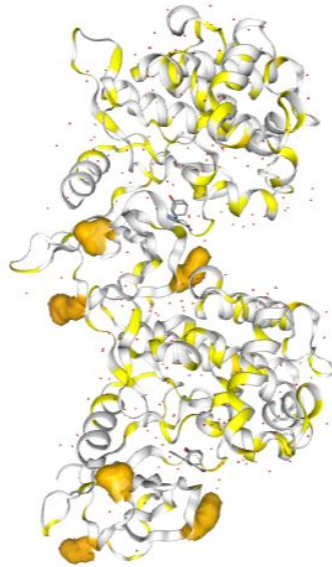


Figura 3.2 Figura que muestra una representación visual en 3D de JAK1

Se encontró además que este gen tiene una ubicación intracelular, y que está asociado con “cytokine receptors” y “cytokine-induced signal transduction”. El gen también se expresó en distintos tipos de cáncer de mama. Es interesante notar que además de esto, las mutaciones de este gen no eran comunes en adultos, según un estudio realizado en 2010.

Tras una búsqueda en PubMed logramos encontrar una relación entre citocina (Cytokine) y el cáncer de colon. En el artículo encontramos que los pacientes con enfermedades inflamatorias del intestino tienen un mayor riesgo de desarrollar cáncer de colon; y en el mismo leímos que las inflamaciones son impulsadas por factores solubles, citocinas y quimiocinas, que pueden ser producidas por las propias células tumorales. Las citocinas y quimiocinas inflamatorias promueven el crecimiento de las células tumorales, perturban su diferenciación y apoyan la supervivencia de las células cancerosas.

Por tanto, una mutación en los genes JAK1 podría llegar a ser causa de hacer crecer cáncer de colon. Las mutaciones somáticas de JAK1 ocurren en “T cell acute lymphoblastic leukemia (T-ALL)”; estas mutaciones parecen ser más frecuentes entre los pacientes adultos con T-ALL, pero la frecuencia varía entre los estudios: 8/38 (21.0%; adult; Flex et al., 2008), 1/49 (2.0%; children; Flex et al., 2008), 2/11 (27.3%; adult; Jeong et al., 2008), 4/108 (3.7%; adult; Asnafi et al., 2009)

Este gen puede encontrarse en el desarrollo del cáncer de colon. Pero aunque sea un gen que se encuentre mayormente expresado en la población joven de acuerdo a la base de datos utilizada para el reto, es un gen que puede mutar más frecuentemente en adultos. Por esto no consideramos que este gen sea exclusivo para ninguno de los dos grupos.

Después, nos pusimos a analizar uno de los genes que había sido mayormente expresado en la población adulta, siendo este el gen NFIL3.



Figura 3.1 Vías de señalización del gen NFIL3

Este gen regula diversos procesos biológicos desde el ritmo circadiano hasta la viabilidad celular. Sin embargo, en un artículo de PubMed encontramos que se han identificado una serie de roles para NFIL3 en la transducción de señales inmunológicas, cáncer y envejecimiento.

NFIL3 puede bloquear el reclutamiento de factores de transcripción ricos en ácido prolina (PAR) a genes proapoptóticos en el cáncer de colon. NFIL3 reprime el gen pro-apoptótico, de sólo BH, BGL-GS en células de cáncer de colon. Por esto, una mutación en el gen NFIL3 podría aumentar los riesgos de cáncer de colon. No encontramos casos de mutación de este gen que llegara a causar cáncer en jóvenes, por lo cuál consideramos que la mutación de NFIL3 aumentando los riesgos de cáncer de colon es un caso específico para la población adulta.

Conclusión

Después de haber realizado todo el proceso de la investigación algunas de las recomendaciones que concluimos que podemos contar para la resolución de este tipo de análisis como el que se hizo en esta investigación es que es de suma importancia elegir una base de datos lo más precisa y confiable posible, sin errores de ningún tipo y con los datos acordes a los que necesitamos para llegar a la conclusión a la que queremos llegar, una vez teniendo la base de datos correcta y de haber checado que no tuviera errores hay que hacer un análisis profundo e investigar a partir de ciertos resultados después de haber identificado y separado la información, por ejemplo en esta investigación separamos los genes hasta llegar a identificar los que más afectan según su p-value, una vez tenidos estos genes separados los ingresamos en una página web para investigarlos más a fondo y poder llegar a ciertas conclusiones, porque para eso son este tipo de investigaciones.

Nos dimos cuenta además de que habíamos empleado una gran cantidad de temas y herramientas que se utilizan en este campo de la biología. Es muy interesante ver la gran complejidad que muchos de estos análisis y sistemas pueden llegar a tener. El uso de herramientas estadísticas como el p-value o el q-value en temas de índole biológica es algo que no parece tener mucho sentido al comienzo, pero terminamos este proyecto sintiéndonos satisfechos de nuestro desempeño por todo lo aprendido y de cómo podemos aplicarlo a algo como la lucha contra el cáncer.

Ahora sí, en cuanto al tema central que trata la investigación, podemos concluir la importancia de la genética a la hora de realizar el análisis acerca del cáncer de colon en adultos jóvenes y adultos viejos, podemos observar que hay una notable diferencia en los genes afectados por el cáncer dependiendo de la edad, y otros factores como por ejemplo que la gente con enfermedades inflamatorias tienen mayor riesgo a sufrir este tipo de cáncer, que la mutación del gen JAK 1 aumenta las probabilidades de contraer el cáncer de colon, que los adultos viejos son más susceptibles a contraer este tipo de cáncer, entre otras cosas.

Referencias

1. Yu, C. et al. Identification of key genes and pathways involved in microsatellite instability in colorectal cancer. *Molecular Medicine Reports* (2019). doi:10.3892/mmr.2019.9849
2. Siegel, R., Miller, K. & Jemal, A. Cancer statistics, 2019. *CA: A Cancer Journal for Clinicians* 69, 7-34 (2019).
3. Secretaría de Salud. Panorama Epidemiológico y Estadístico de la Mortalidad por Causas Sujetas a Vigilancia Epidemiológica en México, 2017. México: Secretaría de Salud, julio 2019. En <<http://tiny.cc/roh5nz>>
4. Edwards, B. et al. Annual Report to the Nation on the status of cancer, 1975-2010, featuring prevalence of comorbidity and impact on survival among persons with lung, colorectal, breast, or prostate cancer. *Cancer* 120, 1290-1314 (2013).
5. Siegel, R. et al. Colorectal Cancer Incidence Patterns in the United States, 1974–2013. *JNCI: Journal of the National Cancer Institute* 109, (2017).
6. Fu, J. et al. Young Patients (≤ 35 years old) With Colorectal Cancer Have Worse Outcomes Due to More Advanced Disease. *Medicine* 93, e135 (2014).
7. Alberts, B. *Molecular biology of the cell*. (Garland Science, 2008).
8. Colon Cancer Diagnosis Improved with Machine Learning Approach. *GEN - Genetic Engineering and Biotechnology News* (2020). En <<https://www.genengnews.com/news/colon-cancer-diagnosis-improved-with-machine-learning-approach/>>
9. 10. What Is Cancer?. *National Cancer Institute* (2020). En <<https://www.cancer.gov/about-cancer/understanding/what-is-cancer>>
10. Klampfer, L. Cytokines, Inflammation and Colon Cancer. *Current Cancer Drug Targets* 11, 451-464 (2011).
11. Keniry, M., Dearth, R., Persans, M. & Parsons, R. New Frontiers for the NFIL3 bZIP Transcription Factor in Cancer, Metabolism and Beyond. *Discoveries* 2, e15 (2014).