

Informe adelanto de resultados “Proyecto de regalías Cáncer Mamario – Omicas”

Fecha: Mayo 15 de 2017.

Tema: Análisis de exomas y driver genes

Resultados Previos:

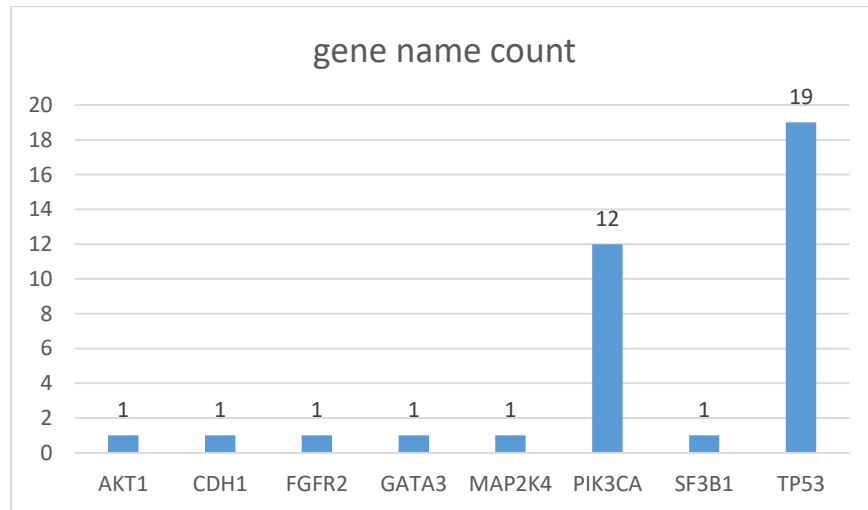
- **DRIVER GENES – análisis previo**

Se extrajo previamente de las bases de datos de: **Cancer Genome Atlas, Broad Genome Institute y Sanger Institute**; un dataset de genes que posiblemente han sido relacionados con diferentes tipos de cáncer alrededor del mundo. De aproximadamente 965 muestras de pacientes, en estas bases de datos se afirma que se han encontrado por lo menos 184 genes involucrados en cáncer mamario. De estos 184 genes solo 37 son candidatos a ser conductores con la más alta probabilidad que ocurra carcinoma de mama.

Se presenta la estadística de elección para el análisis del proyecto OMICAS:

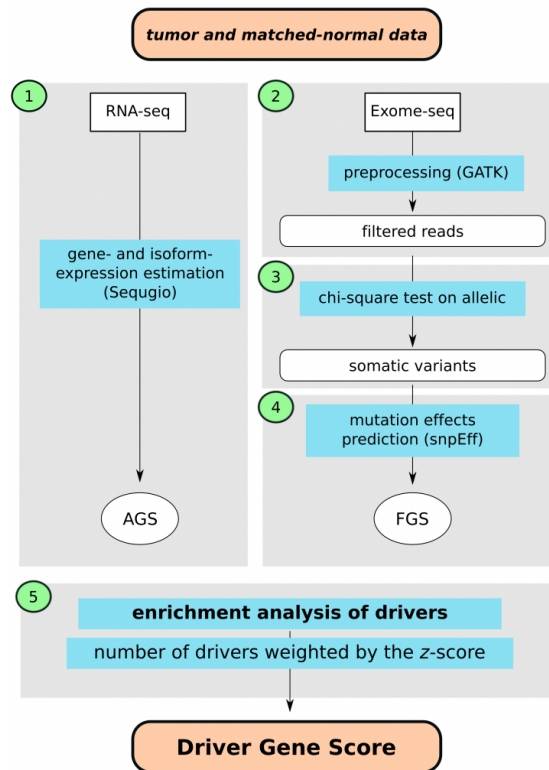
GENE	MUTATION	PERCENTAGE
PIK3CA	G118D	0,35
FGFR2	N550K	0,35
CDH1	SpliceDonorSNV	0,35
MAP2K4	S184L	0,35
TP53	C141Y	0,35
TP53	R342*	0,35
PIK3CA	G1049R	0,35
TP53	R248Q	0,44
TP53	V216M	0,44
TP53	I195T	0,44
TP53	C176F	0,44
PIK3CA	Q546K	0,44
TP53	R306*	0,52
PIK3CA	C420R	0,52
TP53	Y163C	0,52
PIK3CA	E453K	0,52
TP53	H179R	0,52
PIK3CA	Q546R	0,52
TP53	H193R	0,61
TP53	R273H	0,61
TP53	R196*	0,7
TP53	R273C	0,7
GATA3	P409Afs	0,78
TP53	R248W	0,78

Lo que se refiere que existen 8 genes que tienen los mayores records de aparición en los proyectos que involucra cáncer mamario como genes conductores efectivos en una de las mutaciones de su splicing alternativo.



Se propone que empleando la misma técnica, se pueden proponer genes conductores a cáncer mamario específicos para la población colombiana.

El pipeline que se usa para este proceso es el siguiente:



- **Análisis de exomas públicos (pipeline incluido análisis de variantes básico)**

Nombre del proyecto: “Whole exome sequencing suggests much of non BRCA1/BRCA2 familial breast cancer is due to moderate and low penetrance susceptibility alleles”

Código SRA: ERP001755

Numero de muestras: 36 exomas humanos completos.

Plataforma característica: Exome-Seq

CODIGO	ESTADO MUESTRA	NCBI ACCESO
ERR166302	Cáncer mamario	https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=ERR166302
ERR166306	Cáncer mamario	https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=ERR166306
ERR166309	Cáncer mamario	https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=ERR166309
ERR166310	Cáncer mamario	https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=ERR166310
ERR166311	Cáncer mamario	https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=ERR166311
ERR166312	Cáncer mamario	https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=ERR166312
ERR166313	Cáncer mamario	https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=ERR166313
ERR166314	Cáncer mamario	https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=ERR166314
ERR166317	Normal (masculino)	https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=ERR166317
ERR166316	Normal	https://www.ncbi.nlm.nih.gov/sra/?term=ERR166316
ERR166320	Normal (masculino)	https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=ERR166320

ERR166323	Normal	https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=ERR166323
ERR166326	Normal	https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=ERR166326
ERR166327	Normal	https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=ERR166327
ERR166329	Normal	https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=ERR166329
ERR166332	Cáncer mamario	https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=ERR166332
ERR166333	Cáncer mamario	https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=ERR166333
ERR166334	Cáncer mamario	https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=ERR166334

De las 36 muestras disponibles, se han analizado 18 de las cuales 2 corresponden a muestras provenientes de género masculino. Se realiza un análisis previo general de las variantes como primer borrador; los resultados para cada individuo se obtienen de la siguiente forma (ejemplo individuo con código ERR166302):

LOCATION	ERRVCF/ERR166302,Q30F,RAW_VARIANTSHUMAN,VCF
FAILED FILTERS	295
PASSED FILTERS	4148
SNPS	3745
MNPS	0
INSERTIONS	180
DELETIONS	223
INDELS	0
SAME AS REFERENCE	0
SNP TRANSITIONS/TRANSVERSIONS	2,04 (3800/1860)
TOTAL HET/HOM RATIO	0,97 (2039/2109)
SNP HET/HOM RATIO	0,96 (1830/1915)
MNP HET/HOM RATIO	-
INSERTION HET/HOM RATIO	1,07 (93/87)
DELETION HET/HOM RATIO	1,08 (116/107)
INDEL HET/HOM RATIO	-
INSERTION/DELETION RATIO	0,81 (180/223)
INDEL/SNP+MNP RATIO	0,11 (403/3745)

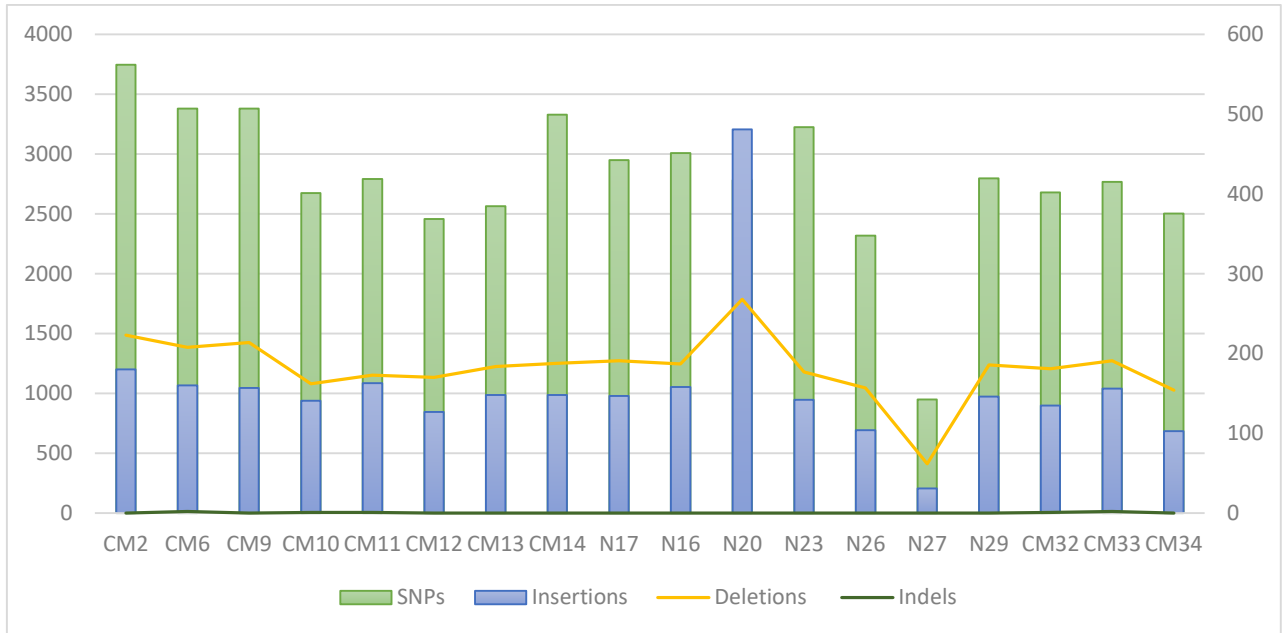
Se realizó una comparación grafica de los individuos como modelo general de análisis, se obtiene lo siguiente:

CODIGO	Código (ID)	ESTADO MUESTRA	SNPs	Insertions	Deletions	Indels
ERR166302	CM2	Cáncer mamario	3745	180	223	0
ERR166306	CM6	Cáncer mamario	3381	160	208	2
ERR166309	CM9	Cáncer mamario	3380	157	214	0
ERR166310	CM10	Cáncer mamario	2674	141	162	1

ERR166311	CM11	Cáncer mamario	2792	163	173	1
ERR166312	CM12	Cáncer mamario	2457	127	170	0
ERR166313	CM13	Cáncer mamario	2565	148	184	0
ERR166314	CM14	Cáncer mamario	3329	148	188	0
ERR166317	N17	Normal	2950	147	191	0
ERR166316	N16	Normal	3007	158	187	0
ERR166320	N20	Normal	2780	481	268	0
ERR166323	N23	Normal	3226	142	177	0
ERR166326	N26	Normal	2318	104	157	0
ERR166327	N27	Normal	950	31	62	0
ERR166329	N29	Normal	2796	146	186	0
ERR166332	CM32	Cáncer mamario	2679	135	181	1
ERR166333	CM33	Cáncer mamario	2769	156	191	2
ERR166334	CM34	Cáncer mamario	2504	103	154	0

A simple viste no hay diferencias significativas entre las muestras; el individuo con código ERR166320 presenta el más alto contenido de inserciones y el individuo con código ERR166327 presenta el más bajo contenido de todas las variantes de estudio con respecto a los demás individuos. No se puede concluir de manera significativa nada ya que este análisis no refiere a variantes específicas; sino a un compendio general como sobre visión del estudio.

Actualmente se está haciendo el llamado de variantes y búsqueda de nuevas para hacer un estudio exhaustivo de cada una de ellas y relacionarlas con los genes conductores a cáncer mamario previamente ya filtrados.



v