IMPLEMENTACIÓN PLATAFORMA EN CIENCIAS OMICAS Y SALUD DEL CANCER MAMARIO, CALI, VALLE DEL CAUCA, OCCIDENTE.

del ***Sistema General de Regalías - SGR***
Fondo Ciencia, Tecnología e Innovación - FCTeI

GOBERNACIÓN VALLE DEL CAUCA

SGR Sistema General de Regalías

DNP Departamento Nacional de Planeación

Universidad del Valle

9 años 1945 - 2015 Conocimiento que transforma vidas

BIOS Centro de Bioinformática y Biología Computacional

bio informática computación

Agenda

1. Árbol del Problema
2. Objetivos
3. Matríz Insumo-Producto
4. Presupuesto según actividades
5. Entidades participantes y aportantes
6. Avances físicos, presupuestales y necesidades

1. Árbol del Problema

CONSECUENCIAS

Difícil acceso al sistema de salud

Alta vulnerabilidad de enfermarse

Altos niveles de deserción laboral y escolar

PROBLEMAS SECUNDARIOS

Deterioro de la calidad de vida

Aumento de costos en el sistema de Salud

Alto riesgo de enfermarse

Aumento de la morbi -mortalidad

EFECTOS

Ausencia de información genómica a nivel poblacional relacionada con el Cáncer mamario en el Sur Occidente (Departamentos del Valle, Nariño, Choco y Cauca)

PROBLEMA CENTRAL

CAUSAS

Inadecuada capacidad de respuesta del sistema de CTeI

Ineficiente sistema de salud pública

Desconocimiento de la constitución genómica de cada persona

Poca acción estatal oportuna

Alto riesgo de la población en regiones vulnerables

Bajo número de investigadores formados en ciencias Omicas y salud

Bajo nivel de financiación para la formación de especialistas

Ineficiente sistema entre oferta y demanda

Ineficiente sistema de vigilancia y control

Baja asignaciones presupuestarias

Baja oportunidades de investigación y difusión

2. Objetivos

Nombre	IMPLEMENTACIÓN PLATAFORMA EN CIENCIAS OMICAS Y SALUD DEL CANCER MAMARIO, CALI, VALLE DEL CAUCA, OCCIDENTE.

Objetivo General

Determinar y analizar la secuencia completa de 321 exomas en 276 individuos con cáncer mamario y 45 individuos control, a fin de identificar genes y variantes genéticas nuevas en el Valle del Cauca, Cauca, Choco y Nariño.

Objetivos Específicos

- **Identificar las personas con Ca mamario y los respectivos controles; y tomar las muestras que cumplan los criterios de inclusión, siguiendo las recomendaciones del Comité de ética, estableciendo un banco de muestras de ADN de pacientes y control.**
- **Determinar las secuencias de los exomas de las muestras obtenidas de las personas involucradas en el estudio.**
- **Analizar e interpretar con métodos bioinformáticos, genéticos, biomédicos y enfoque multifractal las variantes de secuencia en exomas humanos.**
- **Desarrollar una interfaz de Genome Browser para la visualización interactiva de los datos de secuencias de exomas y análisis multifractal de los pacientes y controles.**
- **Diseñar un modelo computacional de un chip para su uso potencial en el diagnóstico del ca mamario.**

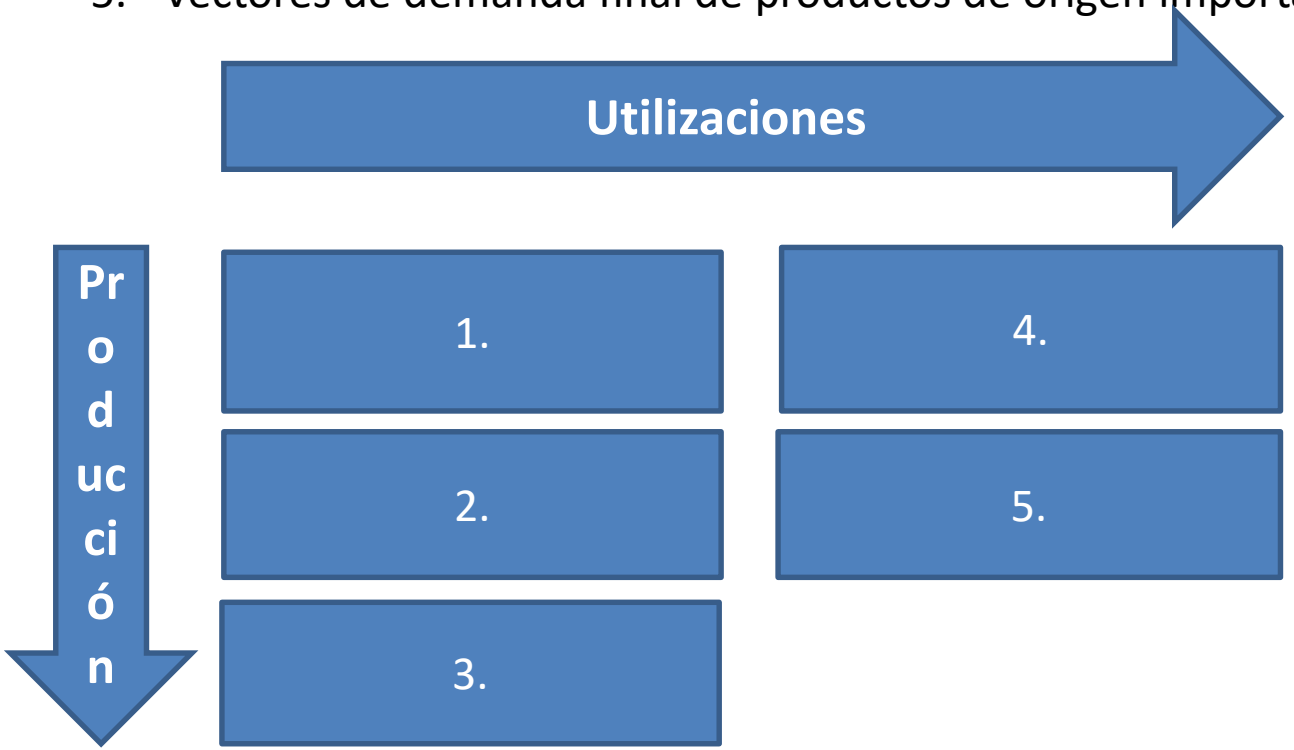
Población a beneficiar (cantidad)	60590
Localización población a beneficiar.	Cabecera Municipal del municipio de Cali, Popayán, Pasto, Quibdo, Tumaco, Guapi, Buena/ra

Entidades Participantes	UNIVERSIDAD DEL VALLE UNIVERSIDAD DEL CAUCA CENTRO BIOINFORMÁTICA
Fase	PREFACTIBILIDAD - FASE 2
Duración	30 meses
Aprobación	N° 027 del 5 de Agosto de 2014
Fecha Inicio	Febrero 2015
Productos entregables:	• Análisis de la información recolectada y establecimiento de un banco de muestras de ADN de pacientes y controles (1 uds). • Plataforma para análisis de secuencias, generación de archivos y datos, genotipificación, **establecimiento de una infraestructura computacional y desarrollo de un software** (1 uds) • Identificación y clasificación de las mutaciones, creación del un portal de internet FTP y HTTP, Análisis multifractal de los exomas y publicación de artículos científicos.(3) (5 uds) • Genome Browser para el estudio de mutaciones de exomas de cáncer mamario en Colombia (1 uds) • **Diagrama de un diseño optimizado para un prototipo de diagnóstico del cáncer de mama tipo chip (1 uds).**

3. Matríz Insumo-Producto

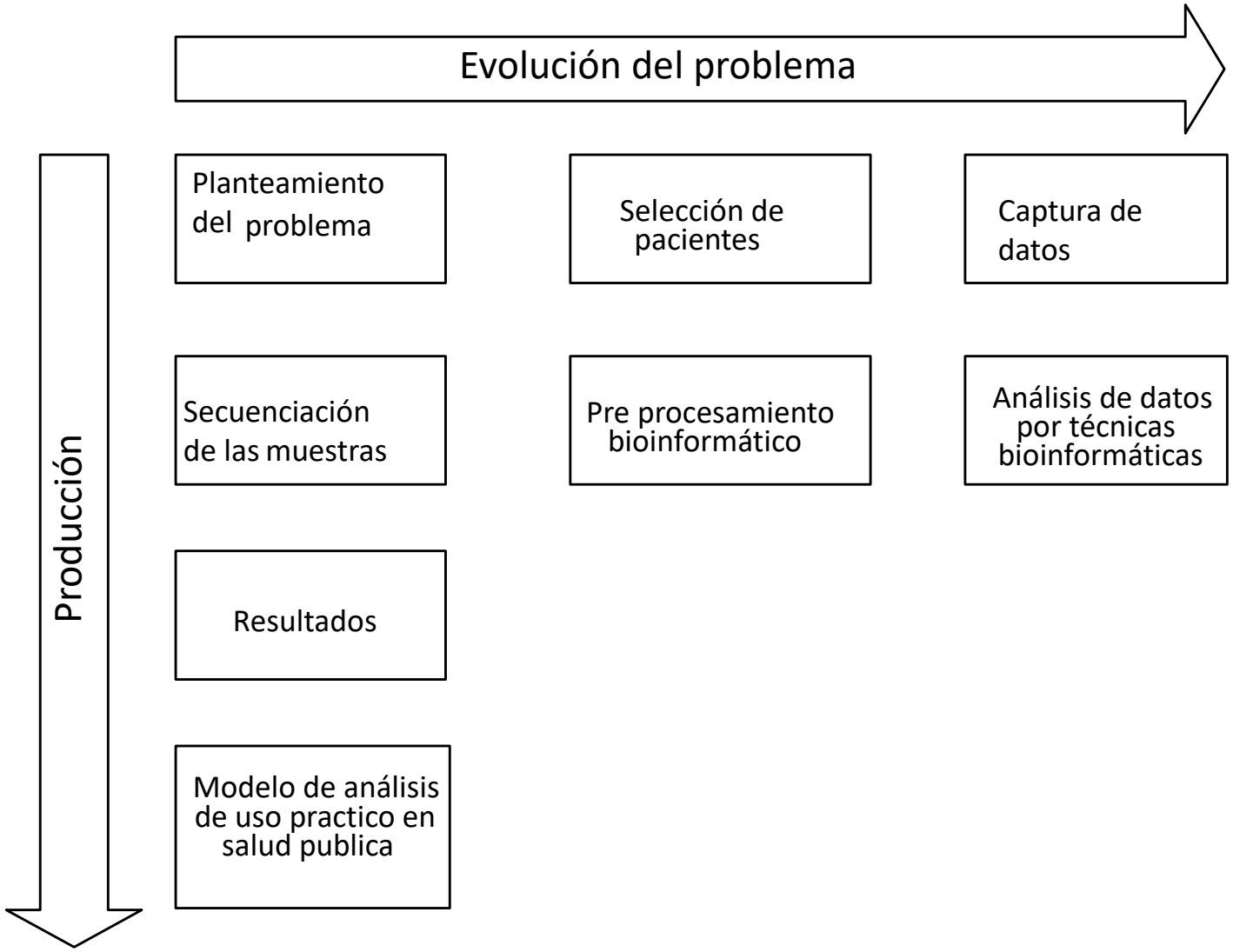
Su estructura está compuesta por cinco submatríces:

1. Matriz de consumos intermedios de origen nacional
2. Vector de consumo intermedio de origen importado.
3. Cuenta de generación del ingreso.
4. Vectores de demanda final de productos de origen nacional
5. Vectores de demanda final de productos de origen importado



El uso de esta matriz no parece el adecuado para nuestro proyecto, ya que esta matriz hace uso de variables econométricas para poder estructurarse, por ejemplo en la metodología dice que uno debe partir de la construcción de un cuadro oferta utilización. Esta matriz es mas para variables económicas entre sectores nacionales e internacionales macroeconómicos que se podrá calcular SI para el proyecto de FACTIBILIDAD en fase III a nivel de diagnóstico.

3. Matríz Insumo-Producto del proyecto



4. Presupuesto según Actividades

AVANCE DEL PROYECTO (De acuerdo al Reporte de Gesproy).
Proyecto: OMICAS
Fecha de reporte: Mayo 16 de 2017

Porcentaje de Avance de Ejecución física con corte al mes Mayo .	61,9%
Porcentaje de Avance de Ejecución financiera con corte al mes Mayo	19,40 %

Producto -MGA	Actividad-MGA	Valor Presupuesto	Valor Ejecutado	Saldo Actividad
Análisis de la información recolectada y establecimiento de un banco de muestras de ADN de pacientes y controles	Dirección general del proyecto	$12.000.000	$7.200.000	$4.800.000
Análisis de la información recolectada y establecimiento de un banco de muestras de ADN de pacientes y controles	Captación de copia de la historia clínica de cada paciente con su diagnóstico y el respectivo consentimiento informado del paciente, familiar y del individuo control.	$14.586.001	$14.586.001	$0,00
Análisis de la información recolectada y establecimiento de un banco de muestras de ADN de pacientes y controles	Obtención de las muestras de tumores de los pacientes y de sangre de los controles sanos, transporte y almacenamiento y banco de tejidos.	$26.724.901	$26.724.901	0,00

Análisis de la información recolectada y establecimiento de un banco de muestras de ADN de pacientes y controles	Procesamiento, extracción y cuantificación de ADN genómico	$85.249.700	$54.987.156	$30.262.544
Análisis de la información recolectada y establecimiento de un banco de muestras de ADN de pacientes y controles	Capacitación en Ciencias Omicas y salud	$13.000.000	$6.600.957	$6.399.043
Plataforma para secuenciación, generación de archivos y datos, genotipificación, establecimiento de una infraestructura computacional y desarrollo de un software	Fragmentación de las secuencias de ADN, preparación de las bibliotecas Illumina enriquecimiento del exoma y generación de agrupamientos. Revisión Bibliográfica.	$29.172.000	$25.581.364	$3.590.636
Plataforma para secuenciación, generación de archivos y datos, genotipificación, establecimiento de una infraestructura computacional y desarrollo de un software	Secuenciación y "basecalling", alineamiento de lecturas (reads) de secuencias y detección de la variación	$711.866.000	$105.429.785	$606.436.215

Plataforma para secuenciación, generación de archivos y datos, genotipificación, establecimiento de una infraestructura computacional y desarrollo de un software	Dirección general del proyecto	$12.000.000	$7.200.000	$4.800.000
Identificación y clasificación de las mutaciones, creación del un portal de internet FTP y HTTP, Análisis multifractal de los exomas y publicación de artículos científicos.	Análisis Exómico y Genómico, anotación de las variantes, identificación de procesos biológicos afectados por las variantes y comparación de genes con bases de datos de exomas de cáncer mamario públicas	$29.171.998	$29.171.998	$0,00
Identificación y clasificación de las mutaciones, creación del un portal de internet FTP y HTTP, Análisis multifractal de los exomas y publicación de artículos científicos.	Aportes en Especie equipos y personal Universidad del Valle, Centro de Bioinformática y Biología Computacional	$1.757.768.170	$1.137.382.221	$620.385.949
Identificación y clasificación de las mutaciones, creación del un portal de internet FTP y HTTP, Análisis multifractal de los exomas y publicación de artículos científicos.	Análisis estadístico y concatenado de exones de cada exoma, análisis multifractal de los exomas.	$96.000.000	$74.637.890	$21.362.110

Identificación y clasificación de las mutaciones, creación del un portal de internet FTP y HTTP, Análisis multifractal de los exomas y publicación de artículos científicos.	Dirección general del proyecto	$12.000.000	$7.200.000	$4.800.000
Identificación y clasificación de las mutaciones, creación del un portal de internet FTP y HTTP, Análisis multifractal de los exomas y publicación de artículos científicos	Capacitaciones, publicaciones y asistencia a eventos	$50.562.000	$14.845.026	$35.716.974
Genome Browser para el estudio de mutaciones de exomas de cáncer mamario en Colombia	Establecimiento de un Genome Browser para desplegar la información de manera comparada con otros repositorios de bases de datos del genoma humano.	$43.731.996	$18.310.656	$25.421.340
Genome Browser para el estudio de mutaciones de exomas de cáncer mamario en Colombia	Dirección general del proyecto	$12.000.000	$7.200.000	$4.800.000

Genome Browser para el estudio de mutaciones de exomas de cáncer mamario en Colombia	Interventoría	$71.406.093	$57.124.872	$14.281.221
Diagrama de un diseño optimizado para un prototipo de diagnóstico del cáncer de mama tipo chip	Diseño del diagrama de un chip potencial para diagnóstico, estudios de tecnologías de microarreglos, factores de error en el procesamiento de los microarreglos, geometría fractal , estudios de simulación con microarreglos y simulación computacional con microarreglos fractales	$138.228.004	$95.852.575	$42.375.429
Diagrama de un diseño optimizado para un prototipo de diagnóstico del cáncer de mama tipo chip	Administración	$129.829.260	$72.239.424	$57.589.836
Diagrama de un diseño optimizado para un prototipo de diagnóstico del cáncer de mama tipo chip	Dirección general del proyecto	$12.000.000	$7.200.000	$4.800.000
	TOTAL	3.257.296.123	$1.769.474.826	1.487.821.297

Producto -MGA	Actividad-MGA	Valor Presupuesto	Valor Ejecutado	Saldo Actividad
Análisis de la información recolectada y establecimiento de un banco de muestras de ADN de pacientes y controles	Dirección general del proyecto	$12.000.000	$7.200.000	$4.800.000
Análisis de la información recolectada y establecimiento de un banco de muestras de ADN de pacientes y controles	Captación de copia de la historia clínica de cada paciente con su diagnóstico y el respectivo consentimiento informado del paciente, familiar y del individuo control.	$14.586.001	$14.586.001	$0,00
Análisis de la información recolectada y establecimiento de un banco de muestras de ADN de pacientes y controles	Obtención de las muestras de tumores de los pacientes y de sangre de los controles sanos, transporte y almacenamiento y banco de tejidos.	$26.724.901	$26.724.901	0,00

Análisis de la información recolectada y establecimiento de un banco de muestras de ADN de pacientes y controles	Procesamiento, extracción y cuantificación de ADN genómico	$85.249.700	$54.987.156	$30.262.544
Análisis de la información recolectada y establecimiento de un banco de muestras de ADN de pacientes y controles	Capacitación en Ciencias Omicas y salud	$13.000.000	$6.600.957	$6.399.043
Plataforma para secuenciación, generación de archivos y datos, genotipificación, establecimiento de una infraestructura computacional y desarrollo de un software	Fragmentación de las secuencias de ADN, preparación de las bibliotecas Illumina enriquecimiento del exoma y generación de agrupamientos. Revisión Bibliográfica.	$29.172.000	$25.581.364	$3.590.636
Plataforma para secuenciación, generación de archivos y datos, genotipificación, establecimiento de una infraestructura computacional y desarrollo de un software	Secuenciación y "basecalling", alineamiento de lecturas (reads) de secuencias y detección de la variación	$711.866.000	$105.429.785	$606.436.215

Plataforma para secuenciación, generación de archivos y datos, genotipificación, establecimiento de una infraestructura computacional y desarrollo de un software	Dirección general del proyecto	$12.000.000	$7.200.000	$4.800.000
Identificación y clasificación de las mutaciones, creación del un portal de internet FTP y HTTP, Análisis multifractal de los exomas y publicación de artículos científicos.	Análisis Exómico y Genómico, anotación de las variantes, identificación de procesos biológicos afectados por las variantes y comparación de genes con bases de datos de exomas de cáncer mamario públicas	$29.171.998	$29.171.998	$0,00
Identificación y clasificación de las mutaciones, creación del un portal de internet FTP y HTTP, Análisis multifractal de los exomas y publicación de artículos científicos.	Aportes en Especie equipos y personal Universidad del Valle, Centro de Bioinformática y Biología Computacional	$1.757.768.170	$0,00	$0,00
Identificación y clasificación de las mutaciones, creación del un portal de internet FTP y HTTP, Análisis multifractal de los exomas y publicación de artículos científicos.	Análisis estadístico y concatenado de exones de cada exoma, análisis multifractal de los exomas.	$96.000.000	$74.637.890	$21.362.110

Identificación y clasificación de las mutaciones, creación del un portal de internet FTP y HTTP, Análisis multifractal de los exomas y publicación de artículos científicos.	Dirección general del proyecto	$12.000.000	$7.200.000	$4.800.000
Identificación y clasificación de las mutaciones, creación del un portal de internet FTP y HTTP, Análisis multifractal de los exomas y publicación de artículos científicos	Capacitaciones, publicaciones y asistencia a eventos	$50.562.000	$14.845.026	$35.716.974
Genome Browser para el estudio de mutaciones de exomas de cáncer mamario en Colombia	Establecimiento de un Genome Browser para desplegar la información de manera comparada con otros repositorios de bases de datos del genoma humano.	$43.731.996	$18.310.656	$25.421.340
Genome Browser para el estudio de mutaciones de exomas de cáncer mamario en Colombia	Dirección general del proyecto	$12.000.000	$7.200.000	$4.800.000

Genome Browser para el estudio de mutaciones de exomas de cáncer mamario en Colombia	Interventoría	$71.406.093	$57.124.872	$14.281.221
Diagrama de un diseño optimizado para un prototipo de diagnóstico del cáncer de mama tipo chip	Diseño del diagrama de un chip potencial para diagnóstico, estudios de tecnologías de microarreglos, factores de error en el procesamiento de los microarreglos, geometría fractal , estudios de simulación con microarreglos y simulación computacional con microarreglos fractales	$138.228.004	$95.852.575	$42.375.429
Diagrama de un diseño optimizado para un prototipo de diagnóstico del cáncer de mama tipo chip	Administración	$129.829.260	$72.239.424	$57.589.836
Diagrama de un diseño optimizado para un prototipo de diagnóstico del cáncer de mama tipo chip	Dirección general del proyecto	$12.000.000	$7.200.000	$4.800.000
	TOTAL	3.257.296.123	$632.092.605	$867.435.348

5. Entidades participantes y Aportantes

Ejecutores

Universidad del Valle

64%

POSTERIS LVMEN MORITVRVS EDAT

1827

UNIVERSIDAD DEL CAUCA

18%

BIOS

CENTRO DE BIOINFORMÁTICA Y BIOLOGÍA COMPUTACIONAL

19%

Colaboradores

Cali

Centro Médico Imbanaco
Vocación de Servicio

HEMATO ONCÓLOGOS S.A
Instituto de Oncología
"Experiencia por la vida"

Popayán

clínica La Estancia
¿Cuánto amor necesita tu vida?

HOSPITAL UNIVERSITARIO SAN JOSE
EMPRESA SOCIAL DEL ESTADO
POPAYAN

Pasto

PATOLOGOS ASOCIADOS

ESE HDSFA
E.S.E HOSPITAL DEPARTAMENTAL SAN FRANCISCO DE ASÍS QUIBDÓ EN LIQUIDACIÓN
NIT. 891.680.047-5

REPÚBLICA DE COLOMBIA

COLCIENCIAS
Departamento Administrativo de Ciencia, Tecnología e Innovación

Grupos de investigación vinculados

Valle del Cauca - UniValle:

- Grupo de Bioinformática (GBioinfo)
- Grupo de Biología Molecular y Patogenicidad (GBMP)
- Grupo de Genética Molecular Humana (GHMH)
- Grupo de investigación de Percepción y Sistemas Inteligentes (GPSI)
- Grupo de Bionanoelectrónica (GBNE
- Grupo de electroquímica

Cauca - UniCauca: Grupo de Biología Molecular, Ambiental y Cáncer, GBIMAC

Caldas: Grupo de Computación Científica -Centro de Bioinformática y Biología Computacional (Manizalez)

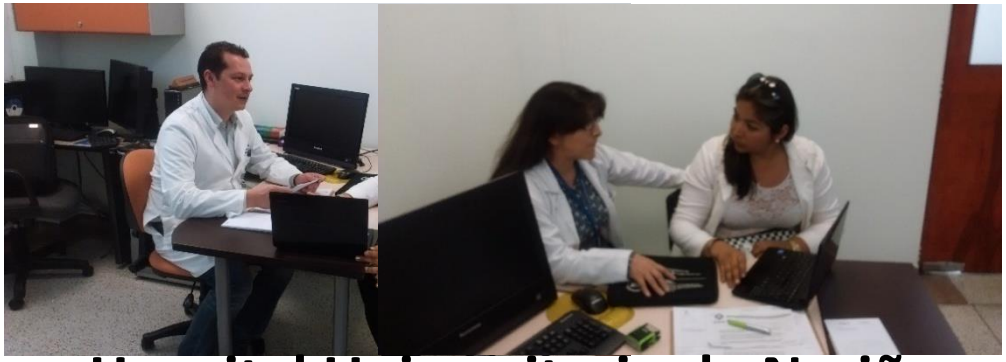
6. Avances Físicos

Clínica Imbanaco

Clínica de Occidente - Cali

Clínica La Estancia- Popayán

Patólogos SA -Pasto



Hospital Universitario de Nariño

6. Avances Físicos: Muestreo

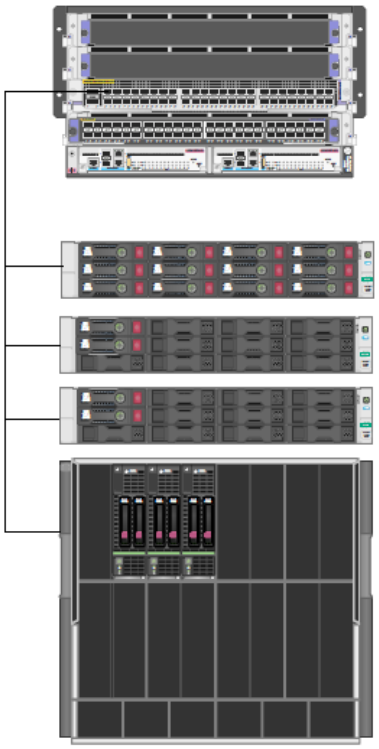
Controles

	# muestras control Parafina	# muestras control tejido	Total	# muestras extracción DNA	% de cumplimiento de acopio de muestras
CAUCA	5	13	18	13	100%
NARIÑO	3	0	3	En proceso	43%
VALLE	0	12	12	12	22%

Casos de cáncer

	# muestras cáncer Parafina	# muestras cáncer Tejido	Total	# muestras extracción DNA	% de cumplimiento de acopio de muestras
CAUCA	9	0	9	En proceso	35%
NARIÑO	10	0	10	En proceso	42%
VALLE	40	30	70	40	22%

6. Avances Físicos: Cluster HPC



Switch core 10Gb con conexión dual entre nodos

Nodo master y Storage 67 Tb de Capacidad, 1 Xeon Phi

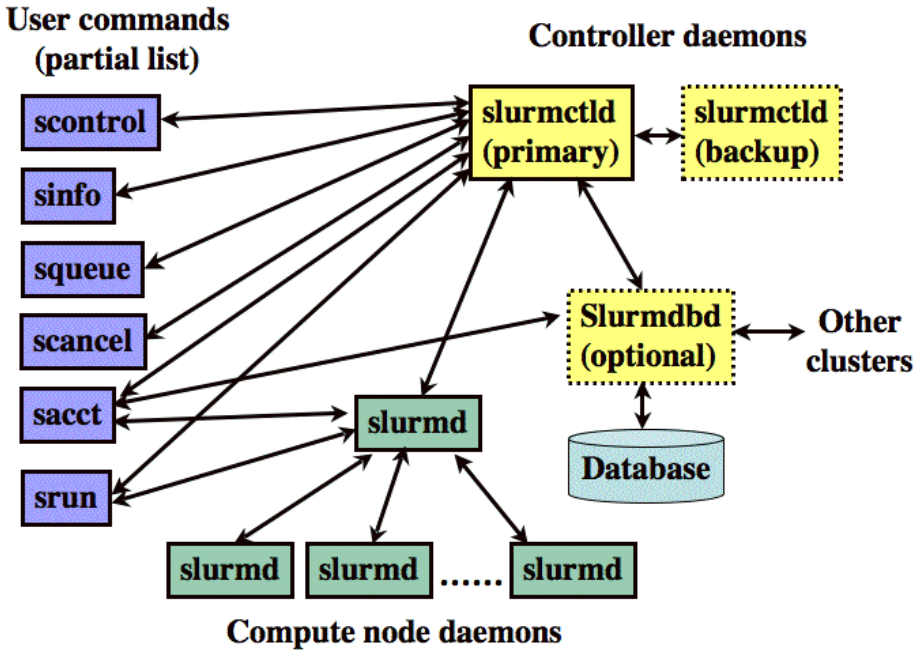
Nodo01 de computo, 2 Xeon phi

Nodo02 de computo, 2 Xeon Phi

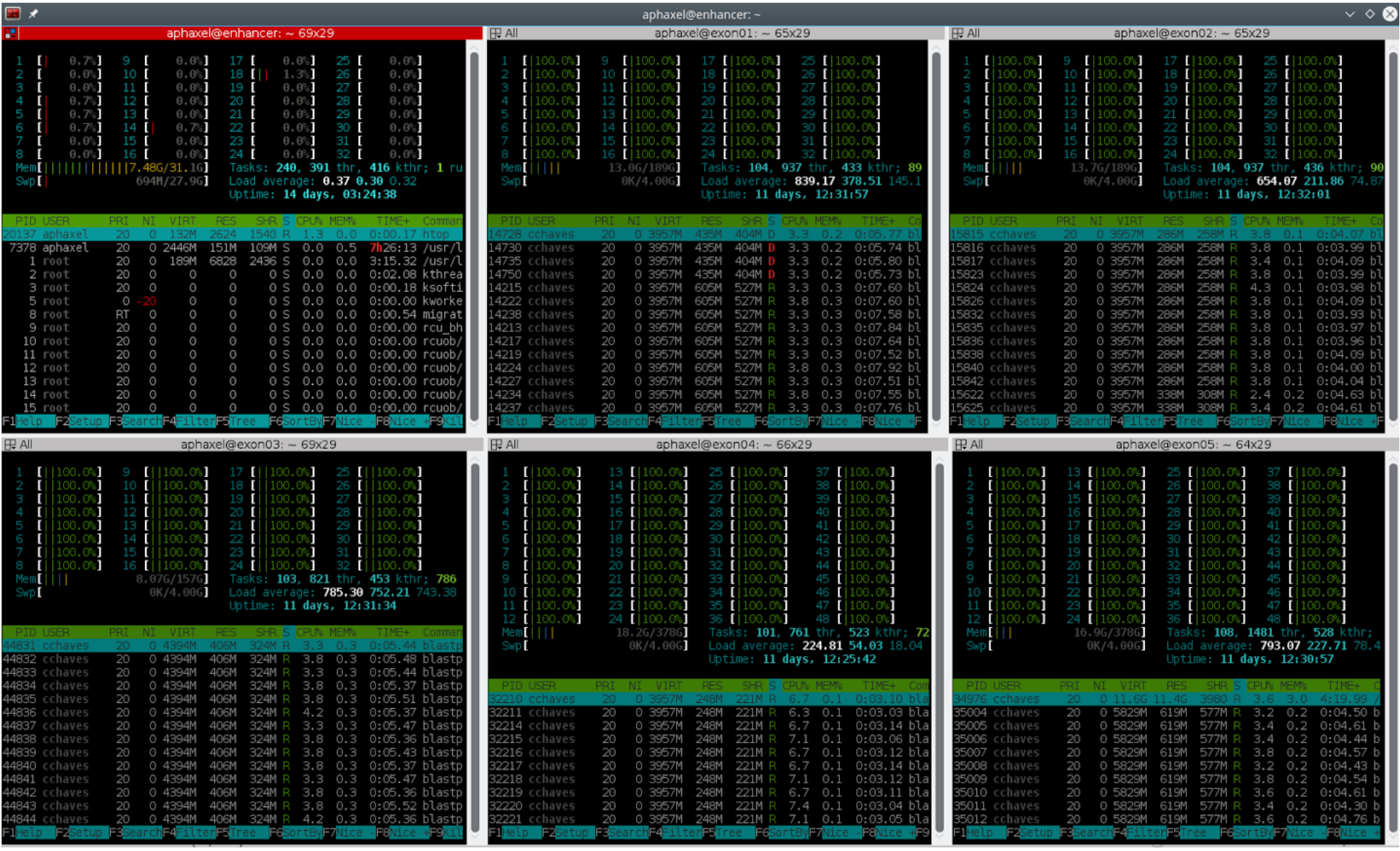
3 nodos tipo blade computo

Arquitectura del clúster HPC, compuesta por 3 nodos de alto desempeño, cada uno con 2 coprocesadores Intel xeon phi de rendimiento teórico de 2.0 TeraFlops, lo cual nos da un pequeño cluster de 6 TeraFlops y un storage (o almacenamiento) de 72 Terabytes.

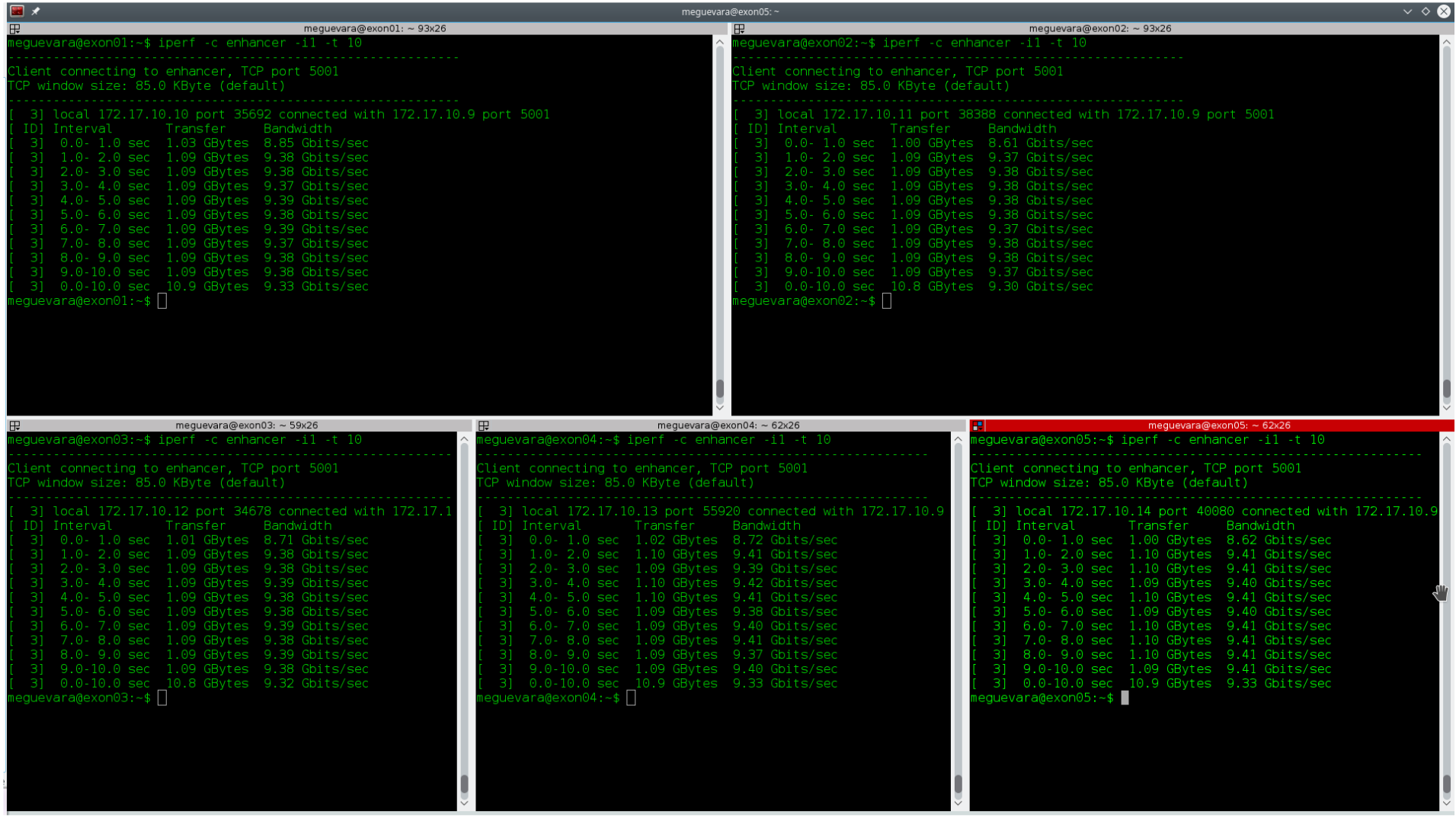
Slurm Arquitectura



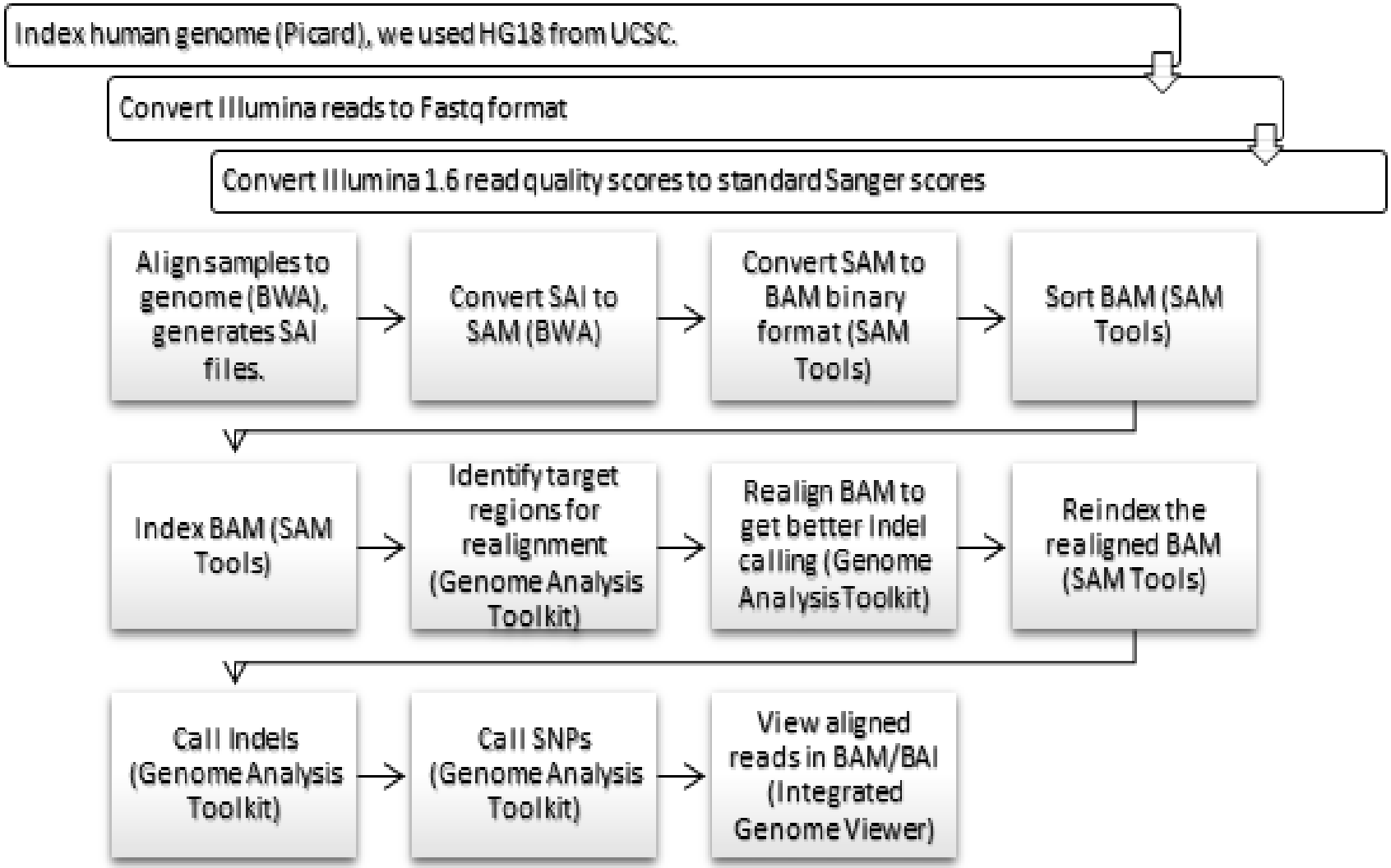
Carga total de uso 100% clúster todos los nodos por 5 días en análisis de secuencias



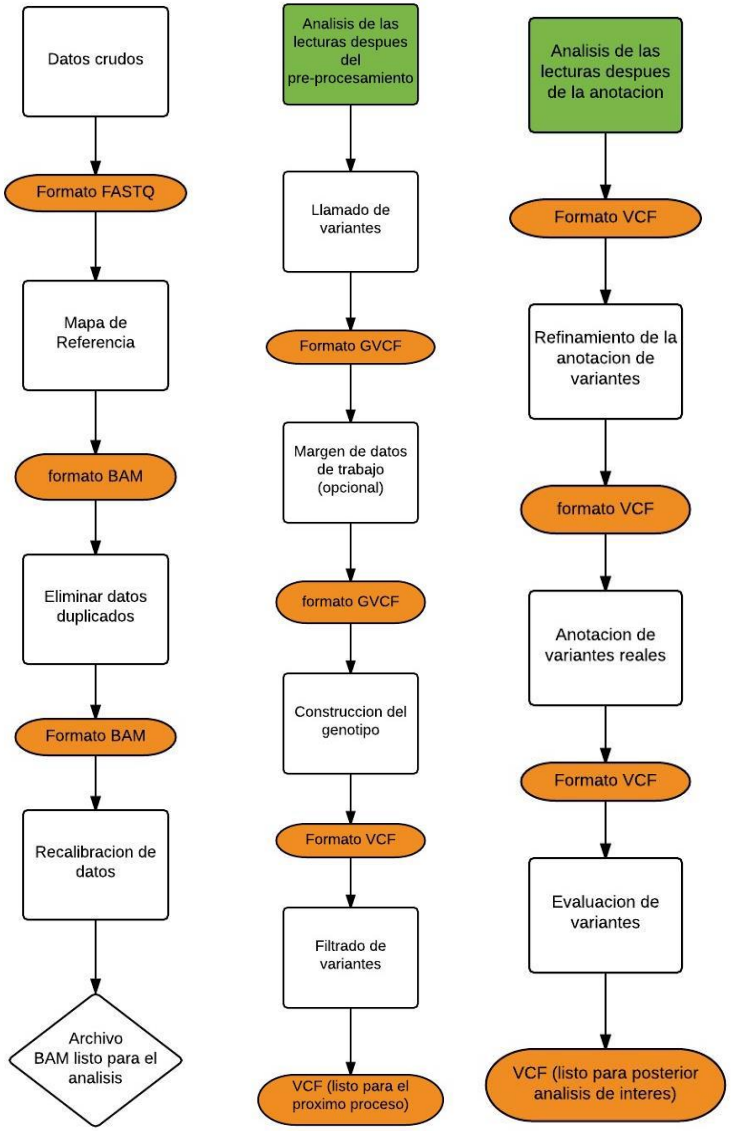
Rendimiento máximo teórico de la red a 10Gb



6. Avances Físicos:
Flujo de trabajo para el análisis de exomas

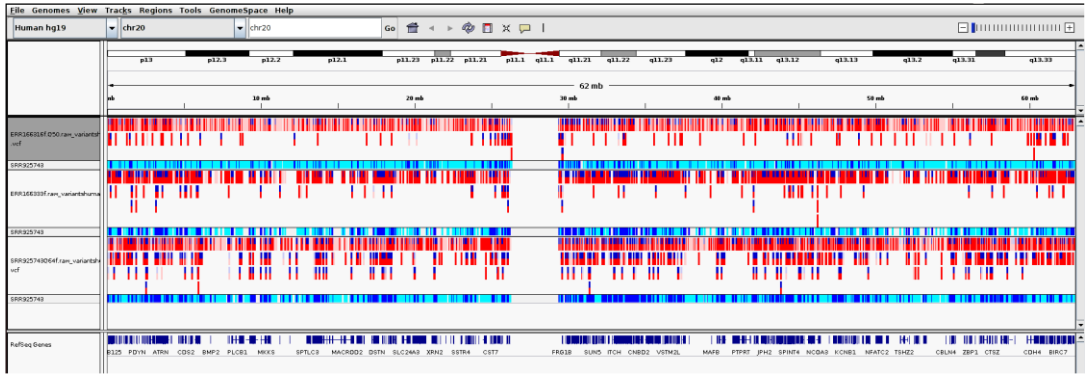


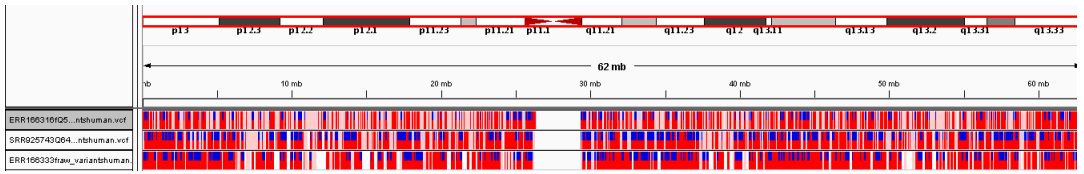
Exome capture alignment & variant calling exercise

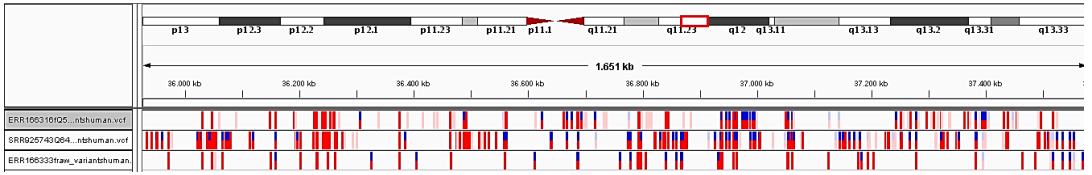


1. Ordenamiento y limpieza de datos
2. Anotación y descubrimiento de variantes
3. Refinamiento de datos y análisis de variantes

CODIGO	ESTADO MUESTRA	NCBI ACCESO
ERR166333	Cáncer mamario	https://www.ncbi.nlm.nih.gov/sra/?term=ERR166333
ERR166316	Normal	https://www.ncbi.nlm.nih.gov/sra/?term=ERR166316
SRR925743	Cáncer mamario	https://www.ncbi.nlm.nih.gov/sra/?term=SRR925743







Diferencia notable de la muestra de cáncer respecto a la normal

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
PIK3CA	E726K	0,78
TP53	R213*	0,87
TP53	Y220C	1,05
TP53	SpliceAcceptorSNV	1,22
SF3B1	K700E	1,39
PIK3CA	N345K	1,48
TP53	SpliceDonorSNV	1,48
PIK3CA	H1047L	1,57
TP53	R175H	1,66
AKT1	E17K	2,35
PIK3CA	E542K	3,4
PIK3CA	E545K	5,31
PIK3CA	H1047R	13,5

Genes conductores

De 965 muestras de pacientes, en bases de datos públicas 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.

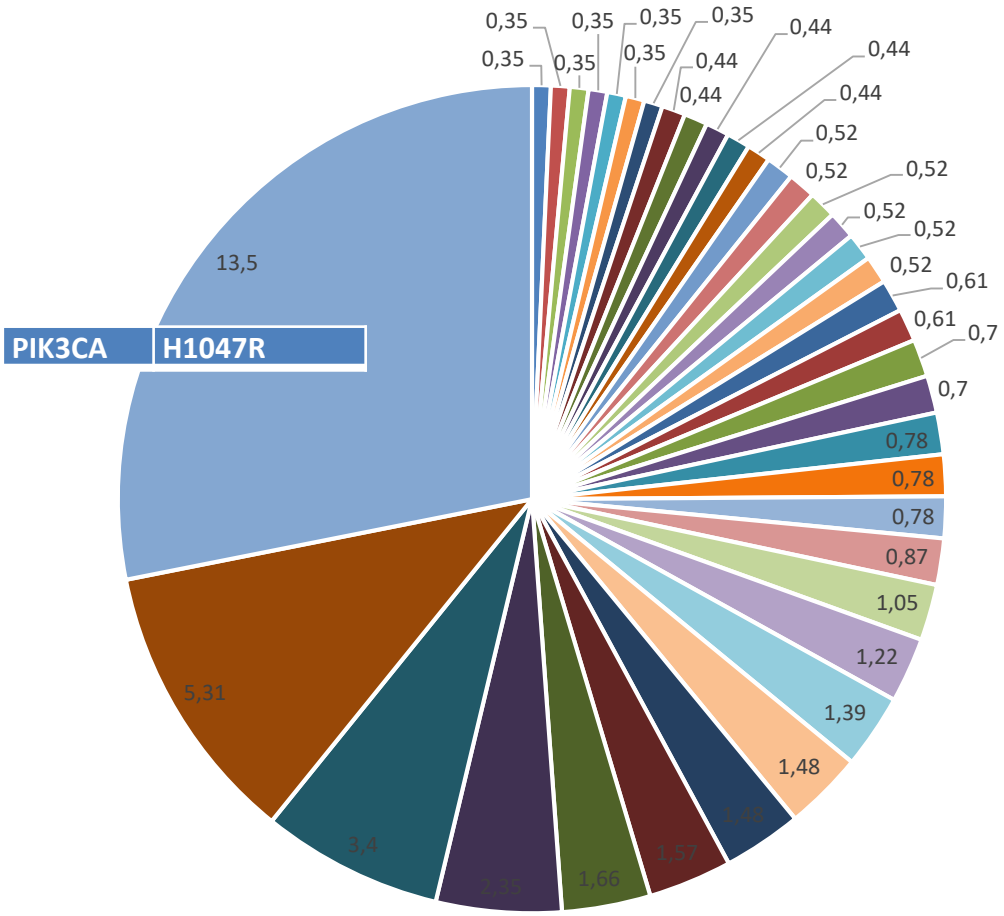
Se presenta la estadística de elección para el análisis del proyecto OMICAS:

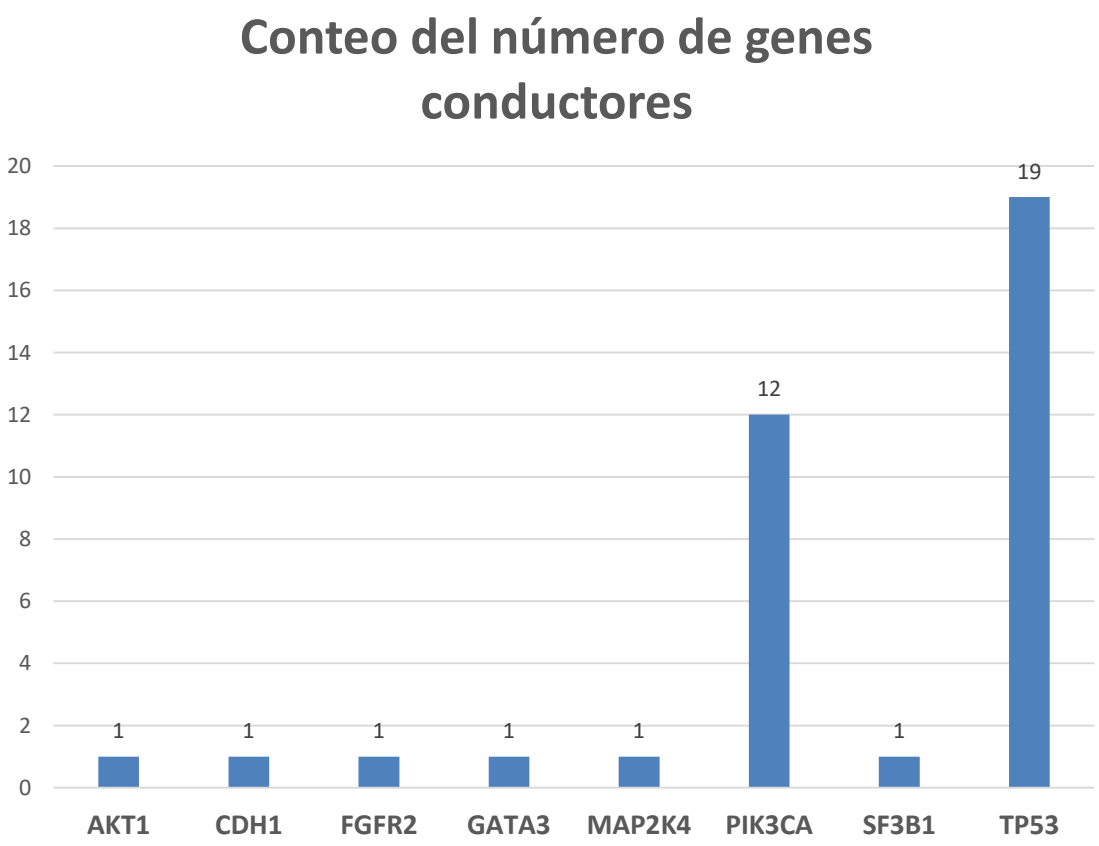
For example:
PIK3CA: phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha.

The ***PIK3CA*** **gene provides instructions for making the p110 alpha (p110**α) protein, which is one piece (subunit) of an enzyme called phosphatidylinositol 3-kinase (PI3K).

These studies suggest that PI3K signaling may be involved in the regulation of several hormones and may play a role in the maturation of fat cells (adipocytes).

PORCENTAJE DE APARICION DE LOS GENES



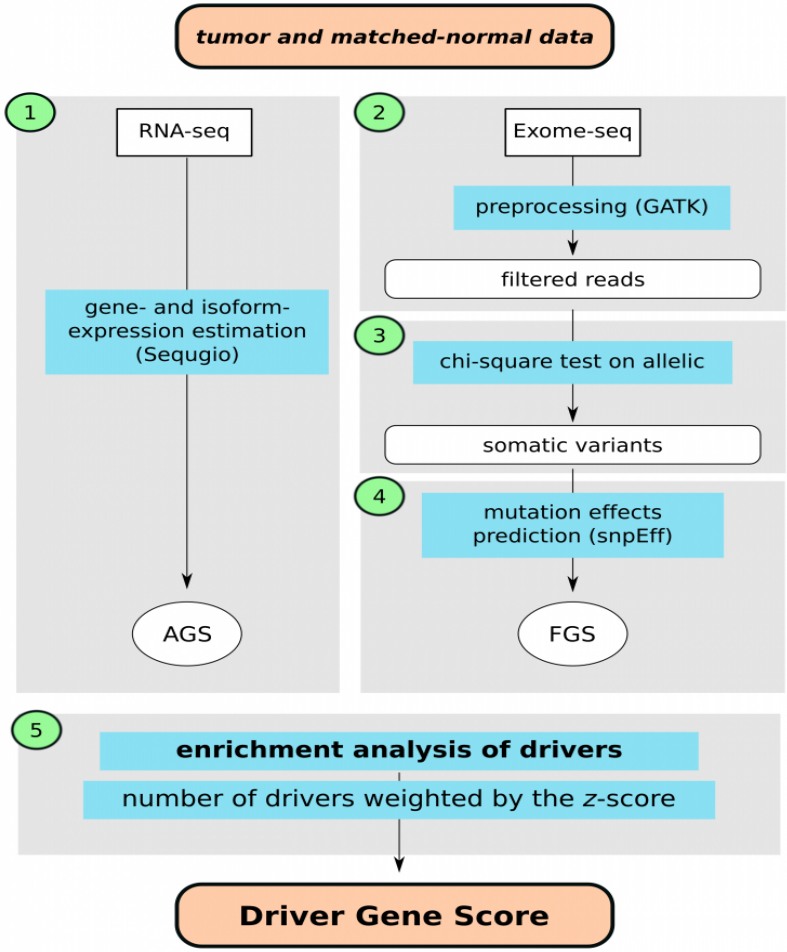


The *TP53* gene provides instructions for making a protein called tumor protein p53 (or p53). This protein acts as a tumor suppressor, which means that it regulates cell division by keeping cells from growing and dividing too fast or in an uncontrolled way.

The p53 protein is located in the nucleus of cells throughout the body, where it attaches (binds) directly to DNA. When the DNA in a cell becomes damaged by agents such as toxic chemicals, radiation, or ultraviolet (UV) rays from sunlight, this protein plays a critical role in determining whether the DNA will be repaired or the damaged cell will self-destruct (undergo apoptosis). **Because p53 is essential for regulating cell division and preventing tumor formation, it has been nicknamed the "guardian of the genome**."

Queremos ver si empleando la misma técnica encontramos genes conductores para el cáncer mamario esporádico específicos para la población del sur occidente colombiano

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



Número de muestras: 36 exomas humanos completos.
Plataforma característica: Exome-Seq
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):

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):

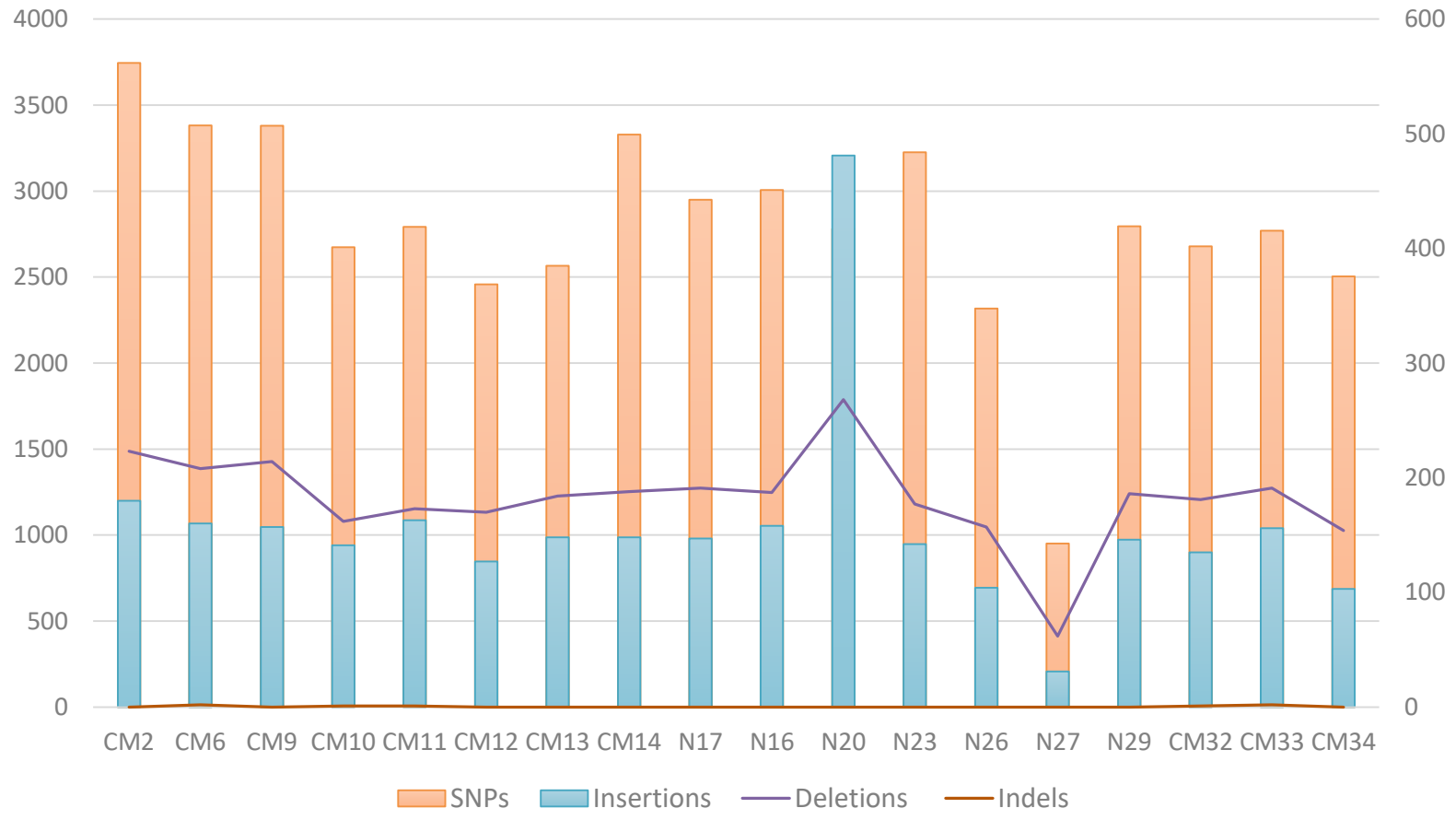
LOCATION	ERRVCF/ERR166302,Q30F,RAW_VARIANTSHUMAN,VCF
FAILED FILTERS	295
PASSED FILTERS	4148
SNPS	3745
MNPS	0
INSERTIONS	180
DELETIONS	223
INDELS	0
SAME AS REFERENCE	0
SNP TRANSITIONS/TRANSVERSIONS	2,04 (3800/1860)
TOTAL HET/HOM RATIO	0,97 (2039/2109)
SNP HET/HOM RATIO	0,96 (1830/1915)
MNP HET/HOM RATIO	-
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.


4000
3500
3000
2500
2000
1500
1000
500
0
600
500
400
300
200
100
0
CM2
CM6
CM9
CM10
CM11
CM12
CM13
CM14
N17
N16
N20
N23
N26
N27
N29
CM32
CM33
CM34
SNPs
Insertions
Deletions
Indels

6. Avances Físicos: Diseño de los planos del prototipo de Microarreglo Microfluídico

Con el propósito de analizar un gran número de muestras en poco tiempo y a bajo costo, se diseñó un prototipo de Microarreglo el cual permite realizar en un solo ensayo la hibridación de 72 muestras con 72 marcadores de cáncer de mama para tener resultados de pronóstico o diagnóstico en pocas horas.

El dispositivo utiliza dos estructuras de microcanales fabricadas en PDMS. Una de las estructuras tiene 72 microcanales en forma radial los cuales se utilizan para inmovilizar los marcadores sobre un substrato de vidrio usando la fuerza centrífuga generada con un dispositivo giratorio. La segunda estructura tiene 72 microcanales dispuestos en forma de espiral los cuales se utilizan para conducir, con la acción de la fuerza centrífuga, todas las muestras simultáneamente hacia los canales con los marcadores. Las muestras que contengan los marcadores hibridizan en los puntos de intersección entre ambos tipos de canales lo cual puede ser detectado posteriormente utilizando un microscopio laser.

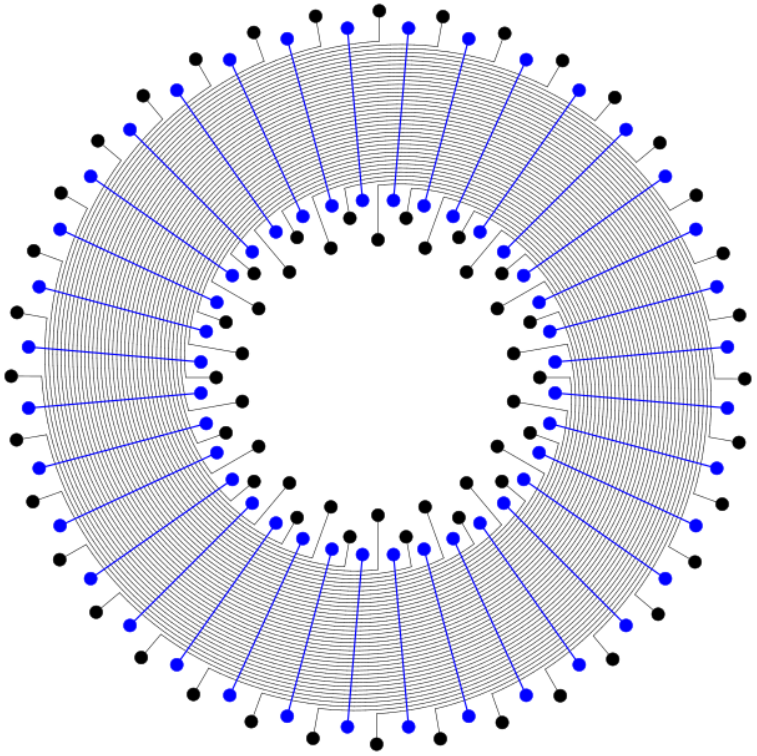


Figura 1: Esquema del Microarreglo Microfluídico (se muestran 36 canales radiales y 36 en espiral para facilitar la visualización).

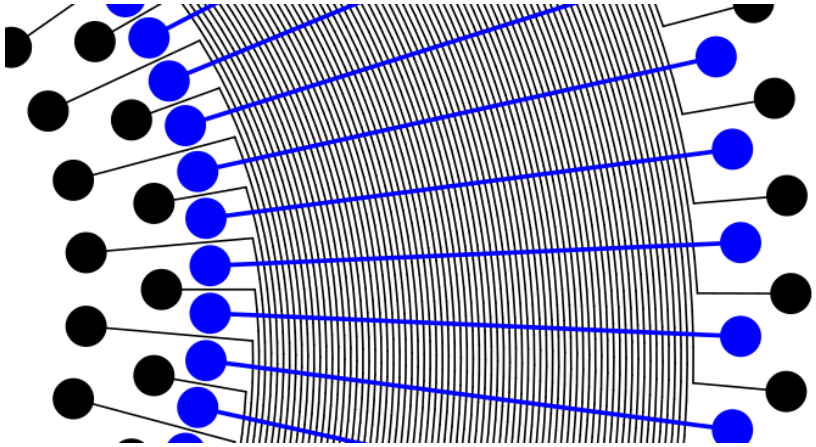
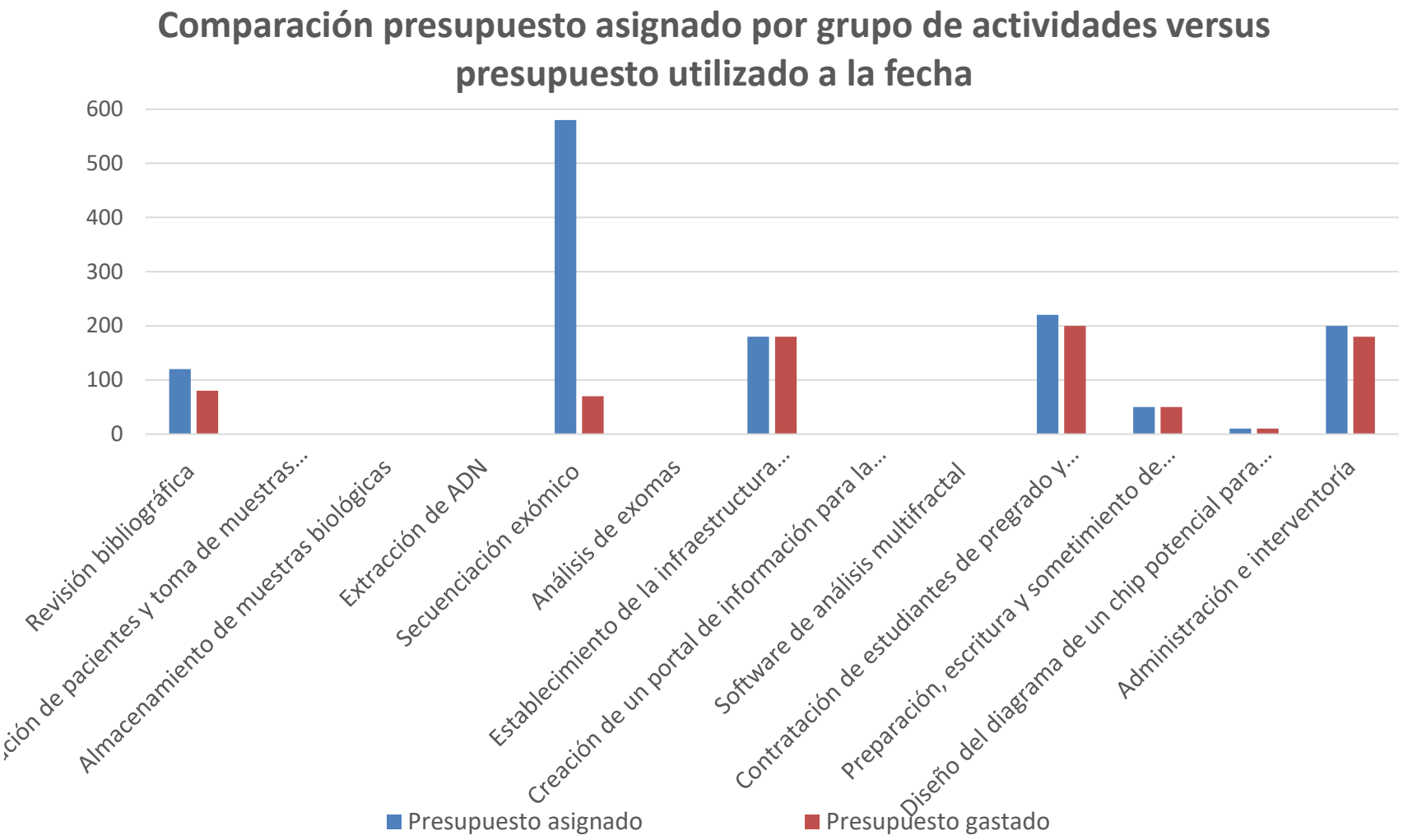


Figura 2: Detalle de la distribución de 72 microcanales radiales y en espiral.

Especificaciones del dispositivo:

Microcanal	Cantidad	Diámetro Total [mm]	Canales		Membrana en PDMS	
			Ancho [µm]	Profundidad [µm]	Diámetro Total [mm]	Espesor [mm]
Espiral	2	89.2	80	30	98	2
Radial	2	93.2	180	30	98	2

6.2. Avances presupuestales



6.3. Necesidades

1. **Aumento del dólar: La TRM**
2. **Intermediarios: FUV**
3. **Las clínicas cobran por análisis de parafinas, entre otras cosas**
4. **Necesidad de mantener el equipo de recurso humano**

Necesidades

Adición presupuestal para terminar el proyecto en Diciembre de 2017

	Actividades	Valor Inicial	Valor Ajustado	Cambio en el valor
1	Captación de copia de la historia clínica de cada paciente con su diagnóstico y el respectivo consentimiento informado del paciente, familiar y del individuo control.	$ 14.586.001	$ 33.124.072	$ 18.538.071
2	Obtención de las muestras de tumores de los pacientes y de sangre de los controles sanos, transporte y almacenamiento y banco de tejidos.	$ 26.724.901	$ 44.088.572	$ 17.363.671
3	Procesamiento, extracción y cuantificación de ADN genómico	$ 85.249.700	$ 98.113.371	$ 12.863.671
4	Análisis estadístico y concatenado de exones de cada exoma, análisis multifractal de los exomas	$ 96.000.000	$ 106.000.000	$ 10.000.000
5	Análisis Exómico y Genómico, anotación de las variantes, identificación de procesos biológicos afectados por las variantes y comparación de genes con bases de datos de exomas de cáncer mamario públicas	$ 29.171.998	$ 63.468.437	$ 34.296.439
6	Secuenciación y basecalling, alineamiento de lecturas (reads) de secuencias y detección de la variación	$ 711.866.000	$ 839.455.156	$ 127.589.156
7	Capacitaciones, publicaciones y asistencia a eventos	$ 50.562.000	$ 80.562.000	$ 30.000.000
8	Interventoría	$ 71.406.093	$ 91.846.769	$ 20.440.676
9	Administración	$ 129.829.260	$ 144.148.416	$ 14.319.156
10	Dirección Gral del proyecto	$ 60.000.000	74.589.160	$ 14.589.160
				$ 300.000.000

Gracias