MACROPROYECTO

Para el Sistema General de Regalías

"IMPLEMETACIÓN PLATAFORMA EN CIENCIAS OMICAS Y SALUD DEL CANCER MAMARIO, CALI, VALLE DEL CAUCA, OCCIDENTE"

Coordinador:

Pedro Antonio Moreno Tovar, Ph.D.

Entidades participantes:

Universidad del Valle
Universidad del Cauca
Centro de Bioinformática y Biología Computacional

Santiago de Cali, junio de 2014

1. INFORMACION GENERAL DE LA PROPUESTA

TÍTULO: PLATAFORMA EN CIENCIAS OMICAS Y SALUD DEL CANCER MAMARIO, CALI, VALLE DEL CAUCA, OCCIDENTE	
Coordinador del Proyecto: Pedro Antonio Moreno Tovar (Biología, Biología molecular, Genómica y Bioinformática)	Universidad del Valle, Facultad de Ingeniería, Escuela de Ingeniería de Sistemas y Computación, Grupo de Bioinformática

Integrantes Ejecutores:

Nombre	**Grupo de Investigación**	**Función**	**Entidad**
Irene Tischer	Bioinformática	Coinvestigador	UniValle
John Sanabria	Estudios Doctorales en Informática - GEDI	Coinvestigador	UniValle
Jaime Velasco	Bionanoelectrónica	Coinvestigador	UniValle
Humberto Loaiza	Percepción y Sistemas Inteligentes	Coinvestigador	UniValle
Guillermo Barreto	Genética Molecular Humana	Coinvestigador	UniValle
Felipe García	Laboratorio de Biología Molecular y Patogenicidad	Coinvestigador	UniValle
Rubén Camargo	Fisicoquímica de Bio y Nanomateriales	Coinvestigador	UniValle
Walter Torres	Electroquimica	Coinvestigador	UniValle
Patricia Eugenia Vélez	Biología Molecular Ambiental y Cáncer (BIMAC). Genética humana y bioética	Coinvestigador	UniCauca
Nancy Janneth Molano Tobar	Biología Molecular Ambiental y Cáncer (BIMAC)	Co-investigador	UniCauca
Maite del Pilar Rada	Biología Molecular Ambiental y Cáncer (BIMAC)	Co-investigador	UniCauca
Néstor Díaz	Biología Molecular Ambiental y Cáncer (BIMAC)	Co-investigador	UniCauca
Siler Amador	Biología Molecular Ambiental y Cáncer (BIMAC)	Co-investigador	UniCauca
Investigador Asociado al proyecto-Ingeniero de soporte	Computación Científica	Co-investigador	Centro de Bioinformática y Biología Computacional

Entidades Colaboradoras:

Coinvestigador	**Especialidad**	**Entidad**	**Ciudad**
Dr. Jaime Rubiano	Oncología, mastología	Hospital Universitario del Valle	Cali
Dr. Rafael Cabal	Patólogo	Particular	Cali
Dra. Marcela Urrego	Oncóloga	Clínica Imbanaco	Cali
Dra. Consuelo Santamaría	Patólogo	Clínica Imbanaco	Cali
Dra. María de los Ángeles Cruz Sánchez	Salud Pública	Secretaria Departamental de Salud del Valle	Cali
Dra. Elsa Patricia Muñoz Laverde	Epidemiología	Escuela de Salud Pública de la UniValle	Cali

Dr. Marino Vélez	Gerencia Servicios de Salud y Salud Publica UniValle	Hospital San Rafael	Zarzal
Dr. Julián Vélez	Estudiante de Administración en Salud, Énfasis Cáncer Mamario	Hospital San Rafael	Zarzal
Jaime López Velazco	Epidemiólogo	Secretaria de Salud del Valle del Cauca	Cali
Wilson Sánchez	Estadístico	Hospital San Rafael Zarzal	Zarzal, Valle
Hernán Vargas	Biomedicina y salud pública	Secretaría de Salud de Bogotá	Bogotá, DC
Marta Domínguez	Biomedicina	Particular	Cali
Fabián Tobar	Bioinformática	Pontificia Universidad Javeriana	Cali
Amparo Acosta	Médico, Genetista	Universidad del Cauca	Popayán
César Edmundo Sarria	Médico y biología molecular	Hospital San José	Popayán
Adriana Chaurra	Química	Universidad del Cauca	Popayán
Rubiel Vargas	Ingeniería de Sistemas	Universidad del Cauca	Popayán
Alí Gómez	Médicina	Hospital Dptal. San Frco. de Asis	Quibdo
Yira E. Gracia	Médicina	Hospital Dptal. San Frco. de Asis	Quibdo
Harold Mauricio Casas	Médicina	Hospital Universitario Dptal. de Nariño	Pasto
King Jordan	Jordan´s lab	Georgia Tech	Atlanta, USA
Fred Vannberg	Vannberg´s lab	Georgia Tech	Atlanta, USA
Alan González	Cirujano plástico	Alan González cirugía plástica & tratamientos no quirúrgicos	Cali
Dago Hernando Bedoya Ortiz	Ingeniero de Sistemas	Centro de Bioinformática y Biología Computacional- CBBC	Manizales

Estudiantes de posgrado y pregrado:

Nombre	**Posgrado/Pregrado (*)**	**Facultad/Entidad**
Christian Arias	Informática	Ingeniería/UniValle– EISC
Carlos Téllez	Informática	Ingeniería/UniValle– EISC
Ivón Bolaños	Biología	Ciencias naturales/UniValle
Mónica Basante	Ingeniería química	Ingeniería/UniValle
Jorge Duarte	Ingeniería electrónica	Ingeniería/UniValle
Carolina Cortés Urrea	Genética	Ciencias básicas de la salud/UniValle
Andrea Coronel	Biomédica	Ciencias básicas de la salud/UniValle
Janeth Molano	Biomédica	Ciencias básicas de la salud/UniValle
Luis Daniel Calderón	Medicina (*)	UniValle
Carlos Fernando Castañeda	Computación	UniCauca
Miguel Guevara	Computación (*)	BIOS
John Delgado	Computación (*)	UniCauca
Diego Martínez	Biología (*)	UniCauca

Marcela Olave	Medicina (*)	UniCauca
Rilder Zambrano	Medicina (*)	UniCauca
Por definir	Ing. de sistemas (*)	UniValle – EISC
Por definir	Ing. de sistemas (*)	UniValle – EISC

Entidades Ejecutoras:
Universidad del Valle, Universidad del Cauca y Centro de Bioinformática y Biología Computacional (Manizales). Tendremos la asesoría internacional de Georgia Tech Institute (USA)

Entidades Colaboradoras:
Hospital Universitario del Valle, Secretaria de Salud Municipal de Cali, Clínica Imbanaco, Secretaria Departamental de Salud de Bogotá, Hospital San José de Popayán, Clínica la Estancia de Popayán, Hospital de Buenaventura, Hospital del Charco, Hospital Dptal. San Frco. de Asis, Hospital Universitario Departamental de Nariño, Hospital San Rafael de Zarzal

Dirección Grupo de Investigación : Pedro Antonio Moreno Tovar	Teléfono: 3148485289	Fax: 3329671
Correo Electrónico: pedro.moreno@correounivalle.edu.co		

Tipo de Proyecto:

Dimensión: Acceso y calidad: Universal y sostenible			
Programa: Protección salud pública			
Fase: II Preinversión, Prefactibilidad			
Investigación Básica (X)	Investigación Aplicada (X)	Desarrollo Tecnológico o Experimental ( )	Otro (¿Cuál?)
Línea de Investigación: Cáncer Mamario, Genómica, Bioinformática, Geometría fractal			
Área del Conocimiento: Biotecnología			
Proyecto interno ( )	Convocatoria VRI ( )	Convocatoria externa (X)	Otra modalidad (¿Cuál?)
Título de la convocatoria: Sistema General de Regalías			
Lugar de Ejecución del Proyecto: Universidad del Valle, Calle 13 No 100-00. - Cali – Valle del Cauca			
Duración del Proyecto (meses): 30 meses			
Fecha de Inicio: Julio de 2014	Fecha estimada de Terminación: Diciembre 2016		
Entidad(es) financiadora(s):	El Sistema General de Regalías		
Nombre de la entidad	Valor	Estado	
Sistema General de Regalías	**$ 1.499´527.953**	T	
Contrapartida Instituciones participantes: Universidad del Valle y Universidad del Cauca			
Efectivo			
Especie:	Universidad del Valle: **$ 1.116'466.050;** Universidad del Cauca: **$ 308´817.000;** Centro de Bioinformática y Biología Computacional: **$ $312.465.120**		
Monto total del Proyecto:	**$ 3.257.296.123**		
Descriptores / Palabras claves: Cáncer de mama, Genómica, Exómica, Multifractales.			

2. RESUMEN EJECUTIVO

PLATAFORMA EN CIENCIAS OMICAS Y SALUD DEL CÁNCER MAMARIO DEL SUR OCCIDENTE (SO)
La presente propuesta de investigación de macro-proyecto en Fase II de pre-inversión, prefactibilidad para su financiación por el Sistema General de Regalías (SGR) en modalidad de investigación básica-aplicada se encuentra enmarcada de acuerdo a la agenda nacional de CTeI, las agendas departamentales del Valle, Cauca, Chocó y Nariño y las municipalidades de Cali, Buenaventura, Popayán, Guapí, Quibdó, Pasto y Tumaco involucradas. Las actividades de ciencia, tecnología e innovación (ACTI) a desarrollar se enmarcan en Investigación y desarrollo experimental.

La plataforma es liderada por la Universidad del Valle e integra la experticia de 7 grupos de investigación, pertenecientes a la Universidad del Valle, uno de la Universidad del Cauca, uno del Centro de Bioinformática y Biología Computacional de Manizales y otro del Dpto. de Biología de Georgia Tech. Institute, USA, como asesor internacional y como tal, la plataforma propende contribuir al desarrollo integral en CTeI de cada uno de los grupos de investigación involucrados y al entendimiento de la génesis del Ca mamario en el SO del país. A su vez, la plataforma se apoya en la colaboración de algunos Hospitales Dptles., Clínicas, Secretaría Dptal. de Salud del Valle, Universidades y de las personas involucradas en el estudio y localizadas en las municipales citadas arriba.

Las ciencias ómicas (*oma*: "conjunto de") o el análisis masivo de información genómica estructural, funcional y comparada es la frontera del conocimiento de la biología integrativa y la medicina genómica (genoma-fenotipo-datos clínicos). La presente propuesta trata de la aplicación de algunos de estos abordajes masivos al sector salud y biotecnológico. En especial, estamos interesados en el estudio de algunas enfermedades Mendelianas, raras y complejas relevantes (y prevalentes) en el SO que aquejan a nuestras gentes. Sin embargo, dado el amplio espectro de estas, nosotros hemos focalizado nuestro primer esfuerzo en el cáncer, y en especial, en el cáncer de mama (o Ca mamario) y la Bioinformática de análisis masivo dada por la experticia en estas áreas de quienes integran la presente propuesta (Vélez, 1991; Moreno et al., 2011; Cifuentes et al., 2013a; Cifuentes et al., 2013b).

Según las estadísticas consignadas en el Plan decenal para el Control del Cáncer en Colombia, 2012-2021 y al Registro Poblacional de Cáncer de Cali, de todos los cánceres, el Ca mamario es la enfermedad prevalente con mayor progresión, con una incidencia por 100.000 habitantes (en 4 años) de 150 casos para Colombia y 190 casos para la ciudad de Cali para personas mayores de 55 años de edad. A pesar de las campañas de prevención, de los mejores instrumentos de diagnóstico, de los diversos programas de detección temprana, de mejores tratamientos y del mayor conocimiento de los factores de riesgos (edad, factores reproductivos, estilo de vida, otros), el Ca mamario sigue aumentando el número de casos alrededor del mundo, especialmente en países occidentales y con incidencia cada vez mayor en personas jóvenes (20 años). El Ca mamario se manifiesta principalmente de dos maneras: esporádico (sin antecedente familiar) en el 85 % de los casos y familiar (con antecedente familiar) en un 5% al 15% de los casos, siendo la manera heredada atribuible a mutaciones por línea germinal de un 5-10% de los casos, y dentro de estos, el 40% se debe a mutaciones en los genes BRCA1 y BRCA2.

Por otra parte, la tendencia en la mortalidad (TAE X 100000 habitantes) en el SO viene en aumento desde 1987 hasta la fecha, lo que ocasiona que un 50% de las defunciones por cáncer de mama correspondan a mujeres del régimen contributivo. Esto implica cargas diferenciales en los años de vida potencial perdidos y una gran carga emocional y afectiva que lesiona las familias. Esto es debido a que el Ca mamario es "una enfermedad de género", afectando principalmente a la mujer, la figura central sobre la cual descansa la unidad familiar, siendo en hombres esta proporción menor del 1%.

Todos estos antecedentes sugieren la existencia de algunos factores genéticos determinantes que desconocemos y que podrían explicar la dicotomía esporádica y familiar. En especial, el origen genético del Ca mamario esporádico (CaME) del cual se conoce muy poco. Por ejemplo, un trabajo reciente hecho mediante tecnologías de secuenciación de última generación, NGS (Next Generation Sequencing), Gracia-Aznarez *et al.* (2013), sugiere que la mayoría de los cánceres mamarios familiar (CaMF) no-BRCA1/BRCA2 podrían ser explicados por la acción de alelos de susceptibilidad de penetrancia baja y/o moderada. Se observa, que a pesar de estos hallazgos, el problema del CaME subsiste.

En Colombia, y en especial en el SO, solamente en la Universidad del Valle se han efectuado algunos estudios que evalúan los factores de riesgo genéticos y las mutaciones relacionadas con el Ca mamario en nuestras gentes (Cifuentes et al., 2013a; 2013b). La siguiente propuesta tiene como objetivo central llevar a cabo un estudio de exomas en 276 individuos afectados con Ca mamario, en modalidad prevalente/transversal y componente analítico caso-control a fin de conocer el estatus de las variantes exómicas relacionadas con los CaME y CaMF en el SO del país. Los resultados esperados de este proyecto impactaran significativamente sobre todas las mujeres del país, dado que la muestra seleccionada es representativa de los principales grupos étnicos que componen el género femenino de nuestro país. En consecuencia, este proyecto, sería la primera ventana y oportunidad para evaluar el estatus exómico del Ca mamario en la región y en Colombia. Esto, con el fin de identificar cual es la contribución de las regiones codificantes del genoma humano a la génesis de estas formas de cáncer.
La identificación de variantes exómicas relacionadas con el desarrollo de las manifestaciones del Ca mamario podría dar luces al entendimiento del origen genético de esta patología. Igualmente, éste análisis permitirá una toma de huellas dactilares más precisa del tumor y contribuir a implementar medidas de diagnósticos ajustadas al exomio de nuestras gentes, de cara hacia la Medicina Personalizada. Por otra parte, dado que la interpretación del genoma es el problema mas crítico a resolver para la praxis médica, nosotros planteamos llevar a cabo un análisis multifractal a fin de evaluar hasta qué grado, dicho análisis nos permitiría contribuir a diferenciar exomas “patológicos” de exomas “saludables”.
A mediano plazo, todos estos estudios pueden tener implicaciones importantes en la predicción de respuesta a la terapeútica y la resistencia y puede conducir al desarrollo de nuevos biomarcadores y terapeútica dirigidas (Zografos et al., 2013).

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:

1) 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.
2) Establecer un banco de muestras de ADN de pacientes y controles.
3) Determinar las secuencias de los exomas de las muestras obtenidas de las personas involucradas en el estudio.
4) Analizar e interpretar por métodos genéticos, bioinformáticos y biomédicos las variantes de secuencia en exomas humanos.
5) Realizar el análisis multifractal de los exomas con inferencias genéticas y biomédicas.
6) 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.
7) Diseñar un modelo computacional de un chip para su uso potencial en el diagnóstico del ca mamario.

Adicional a esto, el proyecto paso por un proceso de MML y se encuentra descrito en formato de MGA. Se publica un video en YouTube: denominado "*Plataforma en ciencias ómicas y salud del cáncer mamario del sur occidente*" (Pendiente).

3. PLANTEAMIENTO DEL PROBLEMA

Introducción

En los últimos años el estudio genético del ser humano ha sido de constante revelación, en cuanto a la explicación de múltiples enfermedades, estudios relacionados con las ciencias Ómicas o el análisis masivo de información genómica estructural, funcional y comparativa, convirtiéndose en la frontera del conocimiento de la biología de sistemas e integrativa. La propuesta se enmarca en la aplicación de estos abordajes masivos al sector salud y biotecnológico. En especial, estamos interesados en el estudio de enfermedades Mendelianas, raras y complejas más relevantes que aquejan a nuestras gentes. Dado el amplio espectro de estas, nosotros focalizaremos el proyecto en el cáncer mamario.

El cáncer, hoy en día, es la pandemia mundial, al igual que algunas enfermedades prevalentes como las cardiovasculares y diabetes, son patologías que han ido incrementado su tasa de mortalidad, como lo reporta Gonzales *et al.* citando a la Agencia Internacional para investigación en Cáncer (IARC) (2006), la cual ha estimado que en el año 2002 hubo 10.9 millones de casos nuevos de cáncer y 6.723.887 muertes por cáncer en todo el globo, la OMS realiza un estimativo mencionando que el número de casos se elevará a 15 millones para el año 2.020, situación que debe ser intervenida desde las diversas áreas de la salud.

3.1 El cáncer mamario en Colombia

Según la OMS, el cáncer de mama es el cáncer más frecuente en las mujeres tanto en los países desarrollados como en los países en desarrollo. La incidencia de cáncer de mama está aumentando en el mundo en desarrollo debido a la mayor esperanza de vida, el aumento de la urbanización y la adopción de modos de vida occidentales. El cáncer de mama es el más común entre las mujeres en todo el mundo, pues representa el 16% de todos los cánceres femeninos. Se estima que en 2004 murieron 519.000 mujeres por cáncer de mama y, aunque este cáncer está considerado como una enfermedad del mundo desarrollado, la mayoría (69%) de las defunciones por esa causa se registran en los países en desarrollo (OMS, Carga Mundial de Morbilidad, 2004).

La incidencia varía mucho en todo el mundo, con tasas normalizadas por edad de hasta 99,4 por 100.000 en América del Norte. Europa oriental, América del Sur, África austral y Asia occidental presentan incidencias moderadas, pero en aumento. La incidencia más baja se da en la mayoría de los países africanos, pero también en ellos se observa un incremento de la incidencia de cáncer de mama. Aunque reducen en cierta medida el riesgo, las estrategias de prevención no pueden eliminar la mayoría de los casos de cáncer de seno que se dan en los países de ingresos bajos y medios (dentro de los que se encuentra Colombia), donde el diagnóstico del problema se hace en fases muy avanzadas. Así pues, la detección precoz con vistas a mejorar el pronóstico y la supervivencia de esos casos sigue siendo la piedra angular del control de cáncer de seno. [1]

A continuación se presenta la incidencia para el período 2002-2006 y por municipios del Valle del Cuaca, Tabla 1.

Tabla 1 > Cáncer de mama de la mujer
Incidencia estimada y mortalidad según departamentos
Colombia, 2002-2006

Departamento	Incidencia estimada			Mortalidad observada		
	Casos anuales	Tasa cruda anual	TAE anual	Muertes en el período	Tasa cruda anual	TAE anual
Antioquia	1.120	39,1	41,5	1.485	10,4	11,0
Atlántico	436	40,2	45,8	563	10,4	11,9
Bogotá	1.386	39,6	43,4	1.820	10,4	11,5
Bolívar	255	27,4	32,6	333	7,1	8,6
Boyacá	136	21,6	21,5	181	5,8	5,6
Caldas	189	38,3	37,5	250	10,1	9,8
Caquetá	32	15,6	22,4	40	3,9	5,8
Cauca	112	18,0	20,8	145	4,7	5,4
Cesar	83	18,6	24,6	102	4,6	6,3
Córdoba	102	14,2	17,6	130	3,6	4,5
Cundinamarca	317	28,3	30,5	421	7,5	8,0
Chocó	24	10,6	13,5	28	2,5	3,3
Huila	142	28,6	33,4	178	7,2	8,5
La Guajira	43	13,0	17,5	52	3,2	4,4
Magdalena	147	26,0	31,6	181	6,4	8,0
Meta	105	27,7	34,6	129	6,8	8,7
Nariño	126	16,6	19,4	162	4,3	5,0
Norte de Santander	178	28,7	33,3	231	7,4	8,7
Quindío	120	44,3	43,5	157	11,6	11,2
Risaralda	191	41,9	42,6	251	11,0	11,1
Santander	358	36,3	38,4	468	9,5	9,9
Sucre	76	20,1	24,0	103	5,5	6,5
Tolima	202	30,0	31,5	270	8,0	8,3
Valle del Cauca	1.039	49,1	51,0	1.352	12,8	13,2
Arauca	21	18,5	27,8	28	4,9	7,7
Casanare	22	15,6	19,7	25	3,5	4,7
Putumayo	10	6,7	11,0	13	1,7	2,9
San Andrés y Providencia	17	48,3	54,8	22	12,5	14,6
Grupo Amazonas*	10	7,2	10,2	12	1,7	2,7
Colombia	6.999	32,6	36,4	9.132	8,5	9,5

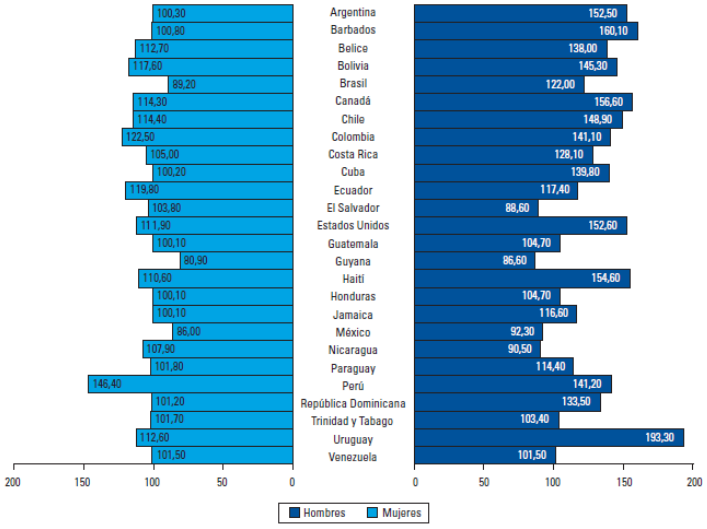
TAE: tasa ajustada por edad.
*Amazonas, Guainía, Guaviare, Vichada y Vaupés

Municipio de residencia	Frecuencia relativa
BUENAVENTURA	1.9%
BUGA	1.1%
CAICEDONIA	1.4%
CALI	69.6%
CALOTO	0.3%
CANDELARIA	1.4%
DAGUA	0.3%
FLORIDA	0.5%
GINEBRA	0.8%
GUACARI	0.3%
JAMUNDI	4.9%
PALMIRA	7.3%
POPAYAN	0.8%
PUERTO TEJADA	0.3%
RESTREPO	0.3%
ROLDANILLO	0.3%
SAN JACINTO DEL CAUCA	0.3%
SANTANDER DE QUILICHAO	1.4%
SEVILLA	0.3%
SONSO	1.1%
TULUA	4.6%
YUMBO	0.8%
ZARZAL	0.3%

FUENTE: INS, Incidencia Estimada y Mortalidad Por Cáncer en Colombia 2002-2006. Frecuencia de la ciudad de residencia de los pacientes con CM en el Valle del Cauca, Fuente propia.

[1] Organización mundial de la Salud, PAHO.

FIGURA 21. Tasas de mortalidad por cáncer (por 100.000 habitantes), según sexo, Región de las Américas, 2002.



Fuente: GLOBOCAN 2002 database, International Agency for Research on Cancer.

Figura 1: Tasa de mortalidad por cáncer, Organización Panamericana de la Salud, PAHO.

Según lo presenta el informe regional Salud en las Américas 2007 (Figura 1), se estimaron 7,6 millones de muertes en el mundo producidas por el cáncer; es decir, 13% de todas las muertes y 21,6% de las muertes por enfermedades crónicas. El cáncer es responsable de 5% de la carga mundial de enfermedad. La mortalidad por cáncer de pulmón y de mama en las mujeres aumentó en la mayoría de los países de América Latina entre 1970 y 2000. Por ejemplo, entre 1970 y 1994, la tasa de mortalidad por cáncer de mama en Costa Rica subió de 6,97 por 100.000 mujeres a 13,42 por 100.000; en Cuba, de 12,33 a 15,82 por 100.000 mujeres; en México, de 4,99 a 9,04, y en Argentina, de 18,64 a 20,99. Por el contrario, en América del Norte, la mortalidad por cáncer de mama ha descendido entre 1985 y 2000. En 2002, la tasa de mortalidad más alta por todos los tumores malignos en los hombres se dio en el Uruguay, con una tasa ajustada por edad de 193,3 por 100.000 (más alta que las de Canadá y los Estados Unidos, que presentaron tasas de 156,6 y 152,6 por100.000 respectivamente, similares a la de Argentina). Entre las mujeres, Perú y Colombia presentaron las tasas de mortalidad ajustadas por edad más altas de América Latina, con 146,4 y 122,5 por 100.000, respectivamente. Esto se evidencia cuando se observan las tasas de años anteriores pues a pesar de que Colombia presentaba la tasa más baja, pasando de 5,2 en 1969 a 9,1 en 1994, la tendencia va en aumento en los últimos años.

Las razones relativamente elevadas entre mortalidad e incidencia en muchos países de América Latina y el Caribe indican que los casos de cáncer de mama no están siendo tratados apropiadamente, por lo que es necesario proporcionar un amplio acceso a los servicios diagnósticos y terapéuticos adecuados. [2]

En 2007, las neoplasias malignas (CIE–10: C00–C99) fueron la causa de 1.130.882 muertes en las Américas, de las cuales 48,4% fueron de mujeres y 51,6% de hombres. En este grupo de causas, las afecciones más importantes fueron las neoplasias malignas de tráquea, de bronquios y de pulmón (CIE–10: C33–C34) con 22,1% del total; de colon, de unión rectosigmoidea, de recto y de ano (CIE–

[2] PAHO, Organización Mundial de la Salud.

10: C18–C21) con 8,7% del total; en las mujeres, fue el cáncer de mama (CIE–10: C50) con 14,8% del total de mujeres; y en los hombres, fueron el cáncer de próstata (CIE–10: C61) con 11,9% del total de hombres, y de estómago (CIE–10: C16) con 5,3% del total. Según los datos de Así Vamos en Salud, en Colombia, para el año 2010 se presento una mortalidad por cáncer de seno de 10,39, siendo la más alta durante los últimos 10 años, Tabla 2.

Tabla 2. Mortalidad por cáncer según primeras causas y sexo, Colombia 2000-2010

Mortalidad por cáncer según primeras causas y sexo, Colombia 2000-2010

Año de defunción	Sexo	Primeras causas de mortalidad por cáncer	Muertes	Tasa cruda por 100.000	TAE por 100.000
2010	Hombres	Tumor maligno de labios, cavidad oral y faringe	258	1,2	1,3
		Tumor maligno de esófago	437	2,0	2,1
		Tumor maligno de estómago	2.796	12,5	13,4
		Tumor maligno de colon, recto y ano	1.261	5,6	6,0
		Tumor maligno de hígado	798	3,6	3,9
		Tumor maligno de páncreas	614	2,7	2,9
		Tumor maligno de tráquea, bronquios y pulmón	2.357	10,5	11,5
		Tumor maligno de próstata	2.431	10,8	11,0
		Tumor maligno de encéfalo y otros de SNC	489	2,2	2,3
		Linfomas no Hodgkin	515	2,3	2,4
		Leucemias	890	4,0	4,1
		Resto de cánceres	3.535	15,8	16,9
	Mujeres	Tumor maligno de estómago	1.709	7,4	6,7
		Tumor maligno de colon, recto y ano	1.456	6,3	5,8
		Tumor maligno de hígado	870	3,8	3,5
		Tumor maligno de vesícula biliar	562	2,4	2,3
		Tumor maligno de páncreas	685	3,0	2,7
		Tumor maligno de tráquea, bronquios y pulmón	1.606	7,0	6,5
		Tumor maligno de mama de la mujer	2.381	10,3	10,0
		Tumor maligno del cuello del útero	1.892	8,2	7,9
		Tumor maligno de ovario u otros anexos	699	3,0	3,0
		Tumor maligno de encéfalo y otros de SNC	443	1,9	1,9
		Leucemias	781	3,4	3,2
		Resto de cánceres	3.985	17,3	16,0

Fuente: Bases de datos Dane
Grupo Vigilancia Epidemiológica del Cáncer, INC

FUENTE: Instituto Nacional de Cancerología, INS.

Por departamentos, Valle del Cauca, San Andrés, Atlántico, Risaralda, Antioquia, Caldas, Tolima, Bogotá, Huila, Quindío, Boyacá y Meta, presentaron tasas superiores a 10 muertes por 100.000 mujeres. A su vez, según los resultados presentados en el atlas de mortalidad por cáncer del Instituto Nacional de Cancerología, el cáncer de seno es la tercera causa de mortalidad por cáncer en mujeres, después del cáncer de cuello uterino y de estómago. La región donde más se concentra la mortalidad por cáncer de seno es la región central del país, seguida de la región nororiental andina y algunos focos en la región Caribe. Los mayores riesgos de muerte coinciden con la ubicación de las capitales departamentales, específicamente en Santa Marta, Barranquilla, Cartagena, Bucaramanga, Medellín, Bogotá y Cali. [3], Tabla 3 y Figura 2.

Tabla 3. Mortalidad por cáncer mamario por departamento, Colombia 2010

[3] Así Vamos en Salud

Mortalidad por Cáncer de Seno por Departamento, Colombia 2010			
Departamento	Mortalidad por Cáncer de Seno	Total de mujeres	Tasa
Amazonas	0	35.437	0,00
Antioquia	370	3.102.432	11,93
Arauca	8	122.624	6,52
Atlántico	154	1.172.798	13,13
Bogotá, D.C.	439	3.815.069	11,51
Bolívar	90	990.125	9,09
Boyacá	66	633.918	10,41
Caldas	59	499.337	11,82
Caquetá	14	221.924	6,31
Casanare	12	159.883	7,51
Cauca	52	650.781	7,99
Cesar	33	483.870	6,82
Chocó	10	239.050	4,18
Córdoba	58	788.368	7,36
Cundinamarca	108	1.240.511	8,71
Guainía	1	18.417	5,43
Guaviare	1	49.054	2,04
Huila	60	539.346	11,12
La Guajira	20	413.469	4,84
Magdalena	38	595.108	6,39
Meta	44	433.465	10,15
Nariño	46	817.161	5,63
Norte de Santander	60	654.058	9,17
Putumayo	4	160.281	2,50
Quindio	31	279.910	11,07
Risaralda	61	474.218	12,86
San Andrés y providencia	5	36.847	13,57
Santander	100	1.017.421	9,83
Sucre	30	399.802	7,50
Tolima	81	690.196	11,74
Total Nacional	2.394	23.042.924	10,39
Valle del Cauca	329	2.256.484	14,58
Vaupés	0	20.384	0,00
Vichada	2	31.176	6,42

FUENTE: Instituto Nacional de Cancerología, INS.

Sanabria y Muñoz establecen que en Colombia, la tendencia en la mortalidad (TAE X 100000) en el SO viene en aumento desde 1987 hasta la fecha, lo que ocasiona que un 50% de las defunciones por cáncer de mama correspondan a mujeres del régimen contributivo. Esto implica cargas diferenciales en los años de vida potencial perdidos y una gran carga emocional y afectiva que desintegra las familias. Esto es debido a que el cáncer de mama prácticamente es una enfermedad de género, y la mujer, es la figura central sobre la cual gira la unidad de la familia.

Dentro de las estadísticas presentadas por IARC, los reportes de cáncer con mayor prevalencia a nivel mundial son el cáncer de pulmón, con más de un millón de casos nuevos, seguidos por el cáncer de mama, cáncer de colon y cáncer del estómago; estos hechos dan a pensar en numerosas variables no solo hereditarias y genéticas que para el cáncer de mama se ha establecido las múltiples mutaciones en genes supresores tumorales, dentro de los cuales se encuentran los denominados BRCA1 y BRCA 2.

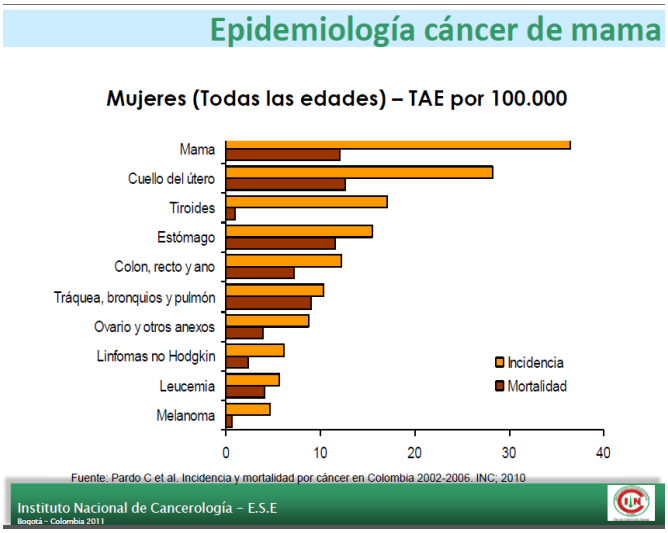


Figura 2. El Ca mamario es el cáncer con mayor incidencia y mortalidad en Colombia.

El Ca mamario se presenta en dos manifestaciones: 1) **Familiar:** Con antecedentes familiares, pero no atribuibles a genética: 5-15% de los casos. Tiene genes de alta y baja penetrancia que no explican el origen de la totalidad de estos tipos de cáncer. **Hereditario:** Atribuidos a mutaciones por línea germinal: 5-10% y dentro de estos, el 40% se debe a mutaciones de los genes BRCA1 y BRCA2. 2) **Esporádico:** Sin antecedentes familiares y se da en el 85% de los casos. No hay genes reportados.

Otros factores de riesgo:
Además de los factores genéticos, el Ca mamario también se manifiesta a través de la acción individual o combinada de varios factores de riesgo, como la edad, factores reproductivos, estilo de vida, entre otros, Figura 3.

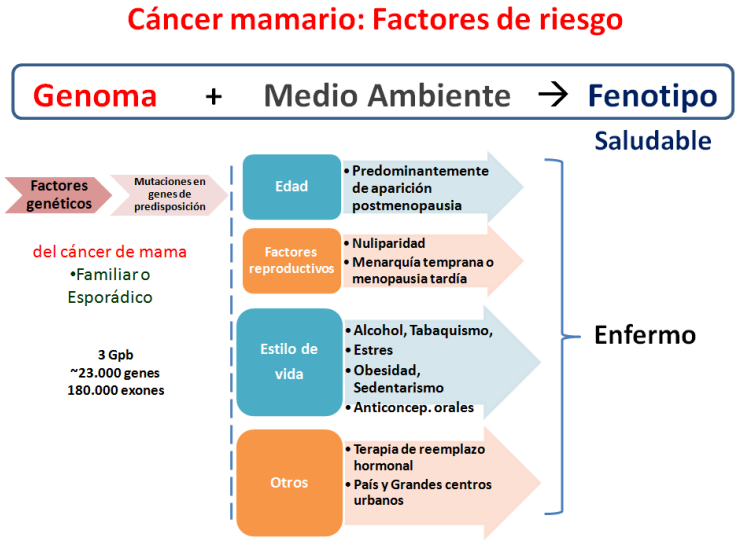


Figura 3. Factores de riesgo para el Ca mamario.

Según el Registro Poblacional de Cáncer en Cali, la tasa de mortalidad (TAE por 100.000) en el Valle del Cauca y su capital, el Ca mamario ha venido en aumento progresivo desde 1987 hasta el

presente. Presentándose una tasa mayor de 168 para la ciudad de Cali en mujeres mayores de 55 años de edad, durante 4 años, Figura 4.

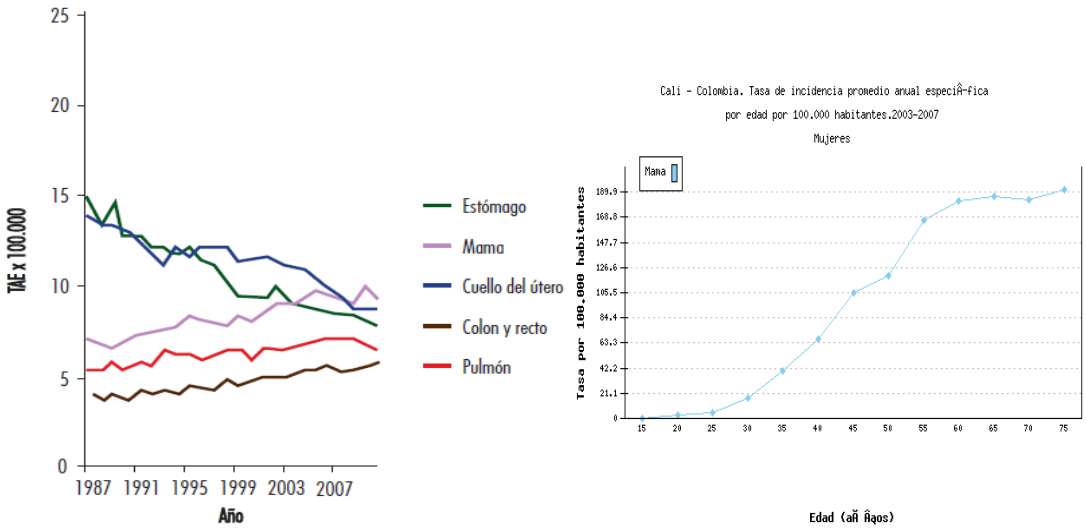


Figura 4. Tasa de mortalidad en la ciudad de Cali.

El diagnóstico clásico del Ca mamario se fundamenta en la detección pre-clínica o clínica del tumor, siguiendo la clasificación de tipos y subtipos TNM. En Colombia, el 90% de las pacientes llegan a la consulta de manera tardía, cuando el tumor ya está en etapa II o más, Figura 5. Todo esto, aumenta los costos al sistema de salud de manera dramática en la medida que la enfermedad progresa en el tiempo. Además, esta progresión afecta de manera emocional la estabilidad de la mujer y de la familia. Si se lograra desarrollar un sistema de diagnóstico temprano efectivo, todos estos sobrecostos emocionales y económicos serían disminuidos sustancialmente.

TIPO DE DIAGNOSTICO	Frecuencia relativa
Etapa: I	10%
Etapa: II	40%
Etapa: III	29%
Etapa: IV	21%
Total	100%

Fuente: Propia

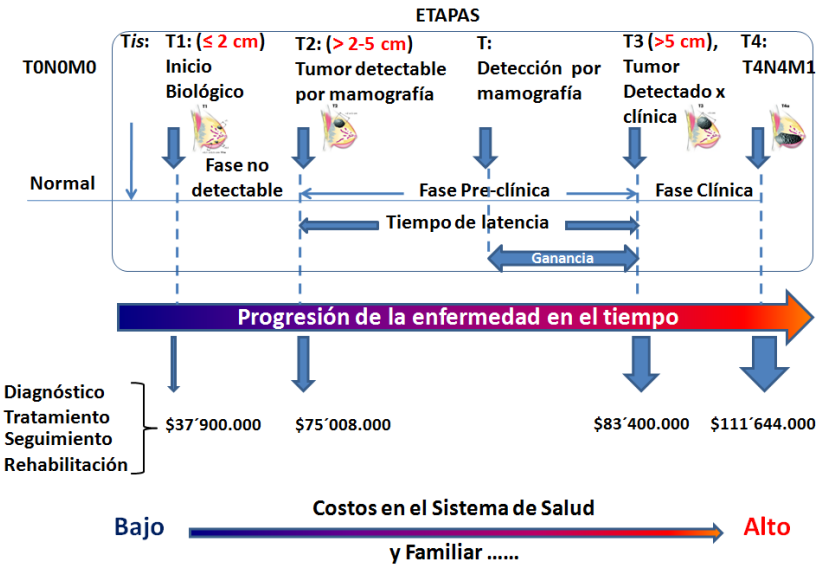


Figura 5. Clasificación del Ca mamario, diagnóstico tardío, progresión y costos.

Otro factor de riesgo menos conocido es el efecto que causa la toma de decisiones e implementación de políticas favorables por parte del Sistema de Salud Colombiano. Sólo hasta hace poco, el INC junto con el Ministerio de Salud han desarrollado una política para diagnosticar de manera masiva la predisposición al Ca mamario con marcadores genéticos aprobados por la OMS. Se espera que esta política contribuya a efectuar manejos de tratamientos más económicos al sistema de salud.

Dado todo esto, por qué es relevante hacer este estudio? A pesar de las campañas de prevención, de mejores instrumentos de diagnóstico, de diversos programas de detección temprana, de mejores tratamientos y de mayor conocimiento de los factores de riesgo, el Ca mamario sigue aumentando el número de casos alrededor del mundo, especialmente en países occidentales. Cada vez mayor incidencia en personas jóvenes (20 años).

3.2 Estudios epidemiológicos descriptivos y prevalentes

Estudios descriptivos

Estos estudios describen la frecuencia y las características más importantes de un problema de salud. Los datos proporcionados por estos estudios son esenciales para los administradores sanitarios así como para los epidemiólogos y los clínicos. Los primeros podrán identificar los grupos de población más vulnerables y distribuir los recursos según dichas necesidades y para los segundos son el primer paso en la investigación de los determinantes de la enfermedad y la identificación de los factores de riesgo.

Los principales tipos de estudios descriptivos son: los estudios ecológicos, los estudios de series de casos y los transversales o de prevalencia.

Estudios transversales: Este tipo de estudios denominados también de prevalencia, estudian simultáneamente la exposición y la enfermedad en una población bien definida en un momento determinado. Esta medición simultánea no permite conocer la secuencia temporal de los acontecimientos y no es por tanto posible determinar si la exposición precedió a la enfermedad o viceversa.

La realización de este tipo de estudios requiere definir claramente:

a. La población de referencia sobre la que se desea extrapolar los resultados.

b. La población susceptible de ser incluida en nuestra muestra delimitando claramente los que pueden ser incluidos en dicho estudio.

c. La selección y definición de variables por las que se va a caracterizar el proceso.

d. Las escalas de medida a utilizar.

e. La definición de "caso".

Los estudios transversales se utilizan fundamentalmente para conocer la prevalencia de una enfermedad o de un factor de riesgo. Esta información es de gran utilidad para valorar el estado de salud de una comunidad y determinar sus necesidades. Así mismo sirven como todos los estudios descriptivos para formular hipótesis etiológicas. Kelsey *et al*., 1986; Hennekens *et al.*, 1987.

En este proyecto se efectuará un estudio descriptivo de prevalencia y transversal con componente analítico caso control.

Tamaño de la muestra: Criterios y variables de las que depende el tamaño de la muestra

1*. El nivel de confianza* (que solemos expresar así: $\alpha = .05$, $\alpha = .01$). Se escogerá un nivel de confianza de .05 (como es práctica común) queremos decir que aceptamos un 5% de

probabilidades de error al rechazar la Hipótesis Nula (de no diferencia). Se trata de minimizar el denominado error Tipo I (aceptamos pocas probabilidades de equivocarnos cuando afirmamos una diferencia o una relación).
2. La *potencia de la prueba*. Por potencia entendemos la probabilidad de no cometer el error denominado Tipo II: *no rechazar la Hipótesis Nula cuando podríamos haberla rechazado*. La probabilidad de cometer este tipo de error se simboliza como β, y la potencia es por lo tanto1-β. Podemos definir la potencia como la probabilidad de rechazar una Hipótesis Nula que es falsa. Así, es razonable establecer una potencia de .80, es decir tener un 80% de probabilidades de detectar una diferencia (o relación).
3. La *magnitud de la diferencia* (o de la relación, etc.) que deseamos detectar y que solemos denominar *tamaño del efecto*. El término *efecto* no implica *causalidad,* sino simplemente el *grado* en que un fenómeno (diferencia, relación, etc.) está presente. Lo *normal* es buscar resultadas (diferencias, relaciones,) *estadísticamente significativos* y no tanto pensar en qué *magnitud* podríamos estar interesados.
4. *Varianza.* Dado que en este diseño, los sujetos son muy iguales dentro de cada grupo, necesitaremos muestras menores para detectar diferencias.

3.3 Estudios genéticos del Ca mamario en Colombia

En Colombia, y en especial en el SO, son pocos los estudios que evalúen los factores genéticos y las mutaciones relacionadas con el cáncer de mama en nuestras comunidades (McCarthy, M/2003). La siguiente propuesta tiene como objetivo central llevar a cabo un estudio de exomas en 276 individuos afectados, en modalidad prevalente/transversal a fin de conocer el estatus de las variantes exómicas relacionadas con los cánceres esporádico y familiar. Esto con el fin de identificar cual es la contribución de las regiones codificantes del genoma humano a la génesis de estos dos formas de cánceres. La identificación de variantes exómicas relacionadas con el desarrollo de estas dos formas de manifestaciones del cáncer de mama podría dar luces a la génesis de esta patología y podría contribuir a implementar medidas de diagnósticos más precisas y ajustadas al genomio de nuestras gentes y de cara hacia la medicina personalizada.

En Colombia los estudios para genes de predisposición familiar a cáncer de mama son escasos, en el 2007 analizaron el gen BRCA1 para 53 familias del centro del país y detectaron cinco mutaciones patogénicas (Torres *et al*., 2007). Se realizó el análisis de 2 mutaciones frecuentes a nivel mundial (185delAG, y 5382insC) en el gen BRCA1 en mujeres con cáncer de mama para el Nor-Oriente del país (Sanabria *et al.,* 2009). En cuanto al Sur Occidente se realizó un estudio en 60 familias en el gen BRCA1, encontrándose un espectro mutacional diferente del encontrado para el centro del país (Cifuentes *et al*., 2010).

El cáncer de mama se considera como la neoplasia femenina más frecuente, constituyéndose en la segunda causa de muerte para las mujeres de todo el mundo. La mayoría de casos de cáncer de mama son esporádicos. No obstante, en aproximadamente el 10% de los afectados se presenta historia familiar positiva, encontrándose generalmente más de un pariente afectado con esta neoplasia, o con otras entidades asociadas como lo es el cáncer de ovario, tumores cuya aparición es temprana y generalmente de forma bilateral, tanto para senos como para ovarios. Esta investigación también se enfocará a casos de cáncer de mama familiar donde la predisposición genética puede estar asociada a mutaciones en genes de supresión tumoral de alta penetrancia como son BRCA1 y BRCA2, o a variantes en genes de baja penetrancia como son la T241M en el gen XRCC3 o la G135C en RAD51.

Para BRCA1 y BRCA2 se han realizado numerosos estudios en diferentes países registrándose una amplia gama de mutaciones y su asociación o no con polimorfismos en genes de baja penetrancia comprometidos con la aparición de cáncer de mama y/u ovario familiar. Un hallazgo interesante para los genes de alta penetrancia es el reporte de alteraciones específicas para cada grupo poblacional las cuales han sido asociadas con el origen geográfico y/o étnico de los individuos muestreados, sugiriendo por tanto que cada población tiene su propia "colección de mutaciones". Estudios previos de nuestro grupo en pacientes con cáncer de mama familiar del Sur-Occidente colombiano y de otros grupos analizando pacientes del centro y sur -oriente colombiano han encontrado diferentes tipos de mutaciones. Es posible que estas diferencias estén relacionadas con el origen geográfico y/o étnico de las muestras o con la sensibilidad de las diferentes metodologías de detección de mutaciones utilizadas.
Teniendo en cuenta que la secuenciación de ADN constituye la metodología de referencia para la detección de mutaciones no solo por su sensibilidad sino también por su capacidad para identificar la naturaleza y tipo de la variante encontrada, en el presente estudio, se hará un barrido de todos los genes (a nivel de exones) del genoma humano con objeto de comparar las mutaciones que se encuentren con las reportadas en la literatura colombiana y mundial. Adicionalmente se definirá si la presencia de variantes en los genes de baja penetrancia XRCC3 y RAD51 están asociadas o no con un riesgo incrementado de desarrollar cáncer de mama y/u ovario en los individuos analizados. Se analizarán muestras procedentes de las diferentes regiones del SO de Colombia para elaborar el mapa mutacional asociado con cáncer de mama y definir si este mapa es diferente de otras regiones en el mundo. De todos los trabajos realizados en Colombia con BRCA sólo el trabajo de UniValle ha secuenciado todo el gen, y es el primero en evaluar asociaciones en genes de baja penetrancia. Los trabajos anteriores buscan las más frecuentes en otras poblaciones.

3.4 Diagnóstico molecular del Ca mamario en Colombia

Actualmente se conocen 33 genes no heredados relacionados con el Ca mamario (Vogelstein et al., 2013), de estos la compañía biotecnológica Oncotype™ de la firma *Genomic Health* utiliza 21 genes (16 genes implicados y 5 genes de control para afinamiento de la prueba). Esta prueba en Colombia tiene un costo de aproximadamente 10 millones de pesos, que bien paga el sistema de salud o de manera particular. Esta prueba está reservada para pacientes en etapa temprana de la enfermedad, sin compromiso axilar y RE (+), bajando las quimioterapias en los EE.UU hasta en 56%. Así, con este procedimiento se asegura un manejo más eficaz de los tratamientos, Tabla 4. El Ministerio de salud y protección social, junto al Instituto Nacional de Cancerología planean un programa de estudio en Bogotá con este tipo de pruebas.

Tabla 4. Genes incluidos en la prueba Oncotype

Gen	Grupo		
Ki-67	Proliferación		
STK15			
Survivin			
Cyclin B1			
MYBL2			
ER	Estrogeno	GSTM1	Otros
PR		CD68	
Bcl2		BAG 1	
SCUBE2		Beta-actina	Referencia (no relacionados con el cáncer)
Stromelysin 3		GAPDH	
Cathepsin L2		RPLPO	
GBR7	Invasión	GUS	
HER2		TFRC	

3.5 Tecnologías de secuenciación de nueva generación y secuenciación de exomas

La secuenciación de nueva generación (NGS) son un conjunto de nuevas tecnologías que permiten la obtención de secuencias de ADN con una alta calidad, cuya secuencia se deriva de una única secuencia y no del ensamble de múltiples secuencias idénticas. Esto permite que dentro de las NGS se puedan observar mutaciones puntuales, o que ocurren en fracciones pequeñas de ADN, como ocurre en los tumores "molecularmente heterogéneos" (Secuenciación profunda). NGS pueden ser limitadas solamente a las regiones del genoma que codifican proteína (Secuenciación exomica), en donde los fragmentos que se obtienen de esta secuenciación pueden ser alineados con un genoma referencia, lo que permite la identificación de inversiones deleciones, translocaciones, y otras mutaciones. La secuenciación exómica involucra la obtención de la región codificante de los genes lo que permite determinar afectación en la estructura proteica y de los procesos metabólicos.

El principio de la secuenciación completa del exoma (WES: whole exome sequencing) subyace en la secuencia nucleótido por nucleótido, de todo el juego de exones del genoma humano de un individuo a una profundidad (80X-100X) de la cobertura necesaria para construir una secuencia de consenso con una alta precisión. Esta secuencia de consenso se compara con los estándares y referencias de lo que es normal en la población y el resultado es interpretado por certificados por la junta directiva de laboratorio y los clínicos. Mediante la secuenciación del exoma de un paciente y comparándolo con la secuencia normal de referencia, variaciones en la secuencia de ADN de un individuo pueden ser identificados y relacionados de nuevo a preocupaciones médicas de la persona en un esfuerzo por descubrir la causa del trastorno médico.

El exoma se refiere a la porción del genoma humano que contiene funcionalmente importantes secuencias de ADN que dirigen el cuerpo para producir proteínas esenciales para que el cuerpo funcione correctamente. Estas regiones de ADN se conocen como exones. Hay aproximadamente 189.000 exones en el genoma humano que representa aproximadamente el 3% del genoma. Estos 189.000 exones están dispuestos en unos 22.000 genes. Se sabe que la mayoría de los errores que se producen en las secuencias de ADN que luego conducen a trastornos genéticos se encuentran en los exones. Por lo tanto, se cree que la secuenciación del exoma a ser un método eficaz de analizar el ADN de un paciente para descubrir la causa genética de las enfermedades o discapacidades. Además, el WES incluye una proyección genoma mitocondrial. Las mitocondrias son estructuras dentro de las células que convierten la energía de los alimentos en una forma que las células pueden utilizar. Aunque la mayor parte del ADN está empaquetado en cromosomas en el núcleo, las mitocondrias también tienen una pequeña cantidad de su propio ADN. Este material genético se conoce como ADN mitocondrial o ADNmt. En los seres humanos, el ADN mitocondrial representa una pequeña fracción del total de ADN en las células. Muchas enfermedades genéticas están relacionadas con cambios en particular, los genes mitocondriales.

La WES es ordenada por un coordinador de un proyecto de investigación o un médico y debe ir acompañada de un formulario de consentimiento y la información clínica detallada. En general, el examen se utiliza para hacer investigación científica y/o cuando la historia clínica del paciente y los hallazgos del examen físico sugieren que existe una etiología genética subyacente. En algunos casos, el paciente puede haber tenido una extensa evaluación que consta de múltiples pruebas genéticas, sin la identificación de una etiología. En otros casos, un médico puede optar por pedir la prueba secuenciación completa del exoma temprano en la evaluación del paciente, en un esfuerzo para acelerar un diagnóstico posible.

Para un paciente con síndrome genético con cáncer no diagnosticado, no se requieren muestras de los padres para interpretar los resultados WES de los probandos. La prueba de Secuenciación Exoma entera es una prueba muy compleja que ha sido recientemente desarrollado para la identificación de cambios en el ADN de un paciente que son causales o relacionados con sus preocupaciones médicas. En contraste con las pruebas de secuenciación actuales que analizan un gen o de pequeños grupos de genes relacionados, a su vez, la prueba de secuenciación completa del exoma analizará los exones o regiones codificantes de miles de genes simultáneamente usando técnicas de secuenciación de próxima generación.

3.6 Estudios de exomas del Ca mamario

Dentro de la secuenciación exómica en cáncer de mama se han encontrado importantes hallazgos respecto a la diversidad y variación genética en la población con cáncer de seno, entre estos trabajos se destacan los siguientes.

Banerji *et al.,* 2012 realizaron la secuenciación de exomas completos para 103 sujetos con cáncer de seno y para diversos subtipos, se comparo con tejidos normales. Los resultados mostraron un total de 4985 mutaciones somáticas en la fracción de los genes que codifica proteína y las regiones adyacentes o puntos de splicing. Seis genes fueron identificados como susceptibles a mutaciones recurrentes (CBFT, TP53, PIK3CA, AKT1, GATA3, y MAP3K1). Este estudio confirma el espectro general de re-arreglos definidos en la célula cancerosa. Por su parte, Shah *et al*., 2012 reportó la secuenciación de RNA (RNA seq) para 80 casos y secuenciación genómica y exómica de 65 casos, todos cáncer de seno TN(Triple Negativo). Los resultados mostraron igualmente conjuntos de genes con altas tasas o frecuencia de mutación, entre estos se encuentra: PIK3CA (10.2%), USH2A (Usher syndrome gene, 9.2%), MYO3A (9.2%), PTEN (7.7%), RB1 (7.7%).

Stephens *et al.,* 2012 analizaron 100 genomas de cáncer de seno de varios subtipos. El resultado mostro varias mutaciones no reportadas (AKT2, ARID1B, CASP8, CDKN1B, MAP3K1, MAP3K13, NCOR1, SMARCD1, y TBX3). Este estudio también revelo una frecuencia particular de mutaciones de citosinas en TpC dinucleotidos; además que la edad de diagnóstico está asociada con el número de mutaciones encontradas.

Pfeffer *et al.* 2013 reportaron a partir de estos trabajos, la alta cantidad de mutaciones que fueron identificada en más de un estudio, han permitido generar un concepto general alrededor del paisaje mutacional de cáncer de seno ha surgido:

- El número de mutaciones en cada tumor simple es altamente variable
- Un alto número de mutaciones están asociadas con más de una enfermedad agresiva
- Agrupaciones de mutaciones en genes pertenecen a vías específicas
- Varios genes están mutados con una alta frecuencia.
- Muchas mutaciones son también privadas (private)
- Varios genes mutados muestran polimorfismo que está asociado con susceptibilidad a cáncer de seno en GWAS.

Actualmente existe una iniciativa para el estudio de exomas: NHLBI Exome Sequencing Project.

Queda claro, con todos antecedentes mencionados, que en Colombia y en el SO, no existe a la fecha un estudio masivo que nos muestre el tipo y la proporción de variación de los genes y su relación con el Ca mamario. De ser aprobado esta propuesta, ésta sería, después de Brasil, un trabajo pionero en Colombia en aplicar estas tecnologías al estudio del Ca mamario en sus dos manifestaciones. En consecuencia, el objetivo central del presente trabajo es detectar y caracterizar mediante secuenciación de exomas las mutaciones presentes en una muestra

colombiana de afectados con Ca mamario. De igual forma el estudio permitirá establecer cuáles son las mutaciones que se presentan con mayor frecuencia, al igual que determinar la existencia de mutaciones propias no halladas en otras partes del mundo y las mutaciones compartidas en CaMF y CaME. Estas variantes pueden contribuir a optimizar, a mediano plazo, las dosis de medicamentos contra el cáncer (McLeod, 2013).
La presente propuesta sigue los lineamientos y procedimientos metodológicos de una Matriz de Marco Lógico y este se encuentra relacionado con la Matriz General Ajustada (MGA) en el material anexo a la propuesta.

4. METODOLOGIAS

4.1 Sobrevisión en diagrama de flujo del proyecto y sus metodologías

La propuesta tiene como eje central una columna vertebral que va del diseño epidemiológico hasta el análisis bioinformático de los exomas, Figura 6.

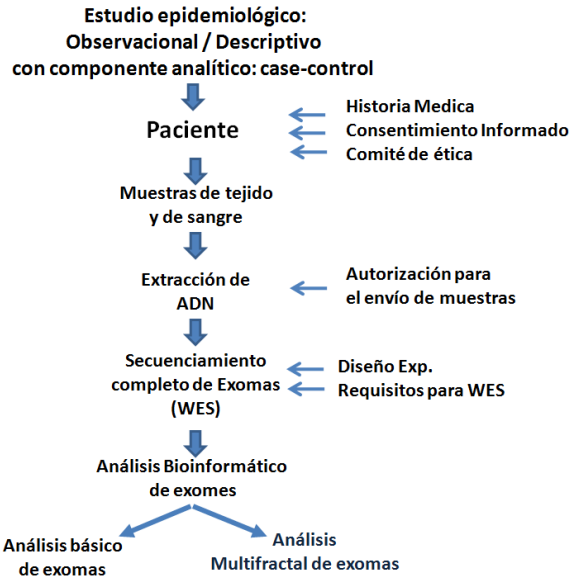


Figura 6. Diagrama de flujo del proyecto y sus metodologías

4.2 Identificar las personas con Ca mamario y los respectivos controles

4.2.1 Pacientes, Controles y Aspectos ético-legales del proyecto

Las personas serán contactadas a través del servicio de salud de los Hospitales departamentales y las clínicas colaboradoras con el estudio, siguiendo los procedimientos y aspectos ético-legales, descritos a continuación.
La investigación está basada en los criterios éticos establecidos en la Resolución 08430 del 4 de Octubre de 1993, en la cual se determinan las normas científicas, técnicas y administrativas en la investigación en el área de la salud. Para el desarrollo del estudio, como en todos aquellos en los que participen seres humanos, es indispensable el respeto a su dignidad y protección a sus derechos, ajustarse a los principios científicos y éticos; deberá siempre prevalecer la seguridad de los participantes y el conocimiento previo de todos los riesgos de la investigación, los cuales estarán plasmados en el consentimiento informado que deben firmar sin excepción alguna (Normas consignadas en el artículo 5, 6 y 9).

Igualmente, se debe proteger la privacidad de los participantes (artículo 8), y será admisible si los beneficios esperados para las comunidades a intervenir sean razonables (artículo 17). Para la investigación en comunidad el investigador debe contar con la aprobación las autoridades de salud, civiles y de los individuos dándoles a conocer la información a que se refieren los artículos 14, 15 y 16 de esta resolución. (Artículo 18). Cuando los individuos que conforman la comunidad no tengan la capacidad para comprender las implicaciones de participar en una investigación, el Comité de Ética en Investigación de la entidad a la que pertenece el Coinvestigador, o de la Entidad en donde se realizará la investigación, podrá autorizar o no que el Consentimiento Informado de los sujetos sea obtenido a través de una persona confiable con autoridad moral sobre la comunidad. En caso de no obtener autorización por parte del Comité de Ética en Investigación, la Investigación no se realizará. Por otra parte, la participación de los individuos será enteramente voluntaria (artículo 19).

4.2.2 Manejo de la confidencialidad de la información

Los participantes (El Coordinador y dos investigadores) de este proyecto tienen clara la responsabilidad y compromiso del manejo confidencial de la información personal a la que se tendrá acceso. Esta estará codificada y no será conocida por terceros, a menos que la persona lo autorice, por ello se solicitara a cada participante de manera libre y voluntaria, la firma de un Consentimiento Informado.

4.2.3 Efectos adversos

Esta investigación no conlleva riesgo alguno porque el análisis es netamente observacional, adicionalmente los procedimientos serán tomados por las entidades y médicos tratantes y las muestras trasferidas a los investigadores para los análisis según protocolos establecidos. Finalmente, el diagnóstico del paciente nunca será puesto en riesgo.

4.2.4 Definición del número de pacientes de la región del SOC

Para el presente estudio se calculó el tamaño de la muestra de la siguiente manera:

A. Población de mujeres (2010. DANE):

Dptos.	**MUJERES**
Choco (Quibdo)	119532
Valle (Cali)	1128101
Cauca (Popayán)	325352
Nariño (Pasto)	408573
Total	1981558

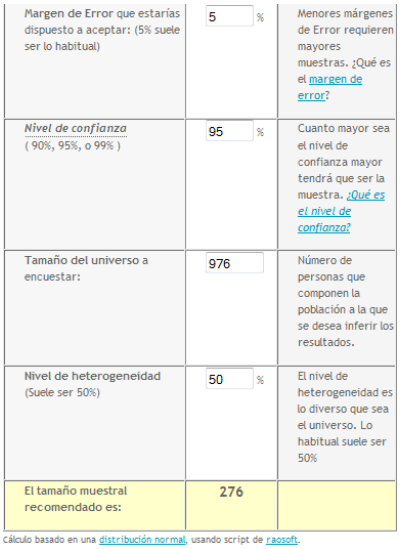
B. Casos anuales con cáncer de mama (INS):

	(2002-2006)
Choco	24
Valle	1039
Cauca	112

Nariño	126
Total	1301

C. Porcentaje de casos con cáncer mamario (T2/T3): 75% = 976 casos

D. Cálculo del tamaño de la muestra con tumor mamario (T2/T3):

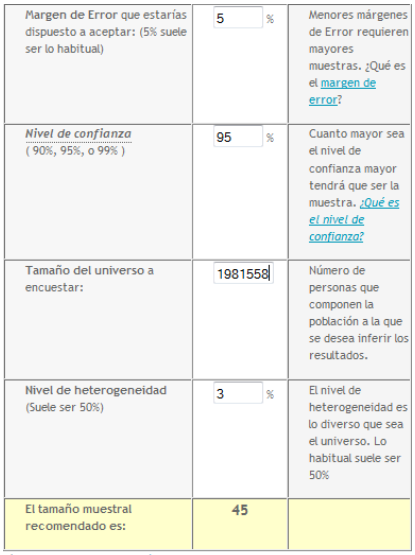


Margen de Error que estarías dispuesto a aceptar: (5% suele ser lo habitual)	5 %	Menores márgenes de Error requieren mayores muestras. ¿Qué es el margen de error?
Nivel de confianza (90%, 95%, o 99%)	95 %	Cuanto mayor sea el nivel de confianza mayor tendrá que ser la muestra. *¿Qué es el nivel de confianza?*
Tamaño del universo a encuestar:	976	Número de personas que componen la población a la que se desea inferir los resultados.
Nivel de heterogeneidad (Suele ser 50%)	50 %	El nivel de heterogeneidad es lo diverso que sea el universo. Lo habitual suele ser 50%
El tamaño muestral recomendado es:	276	

Cálculo basado en una distribución normal, usando script de raosoft.

Dado que no conocemos las frecuencias genotípicas de los genes "conductores" en la población afectada con cáncer de mama, asumimos una heterogeneidad del 50%.

E. Cálculo del tamaño de la muestra control:



Margen de Error que estarías dispuesto a aceptar: (5% suele ser lo habitual)	5 %	Menores márgenes de Error requieren mayores muestras. ¿Qué es el margen de error?
Nivel de confianza (90%, 95%, o 99%)	95 %	Cuanto mayor sea el nivel de confianza mayor tendrá que ser la muestra. *¿Qué es el nivel de confianza?*
Tamaño del universo a encuestar:	1981558	Número de personas que componen la población a la que se desea inferir los resultados.
Nivel de heterogeneidad (Suele ser 50%)	3 %	El nivel de heterogeneidad es lo diverso que sea el universo. Lo habitual suele ser 50%
El tamaño muestral recomendado es:	45	

Cálculo basado en una distribución normal, usando script de raosoft.

Se asume una heterogeneidad del 3%, dado que muy pocos genes "conductores" estarían mutados en la población control.

En conclusión, se analizarán 321 exomas de pacientes y controles. Si los exomas crudos (sin análisis bioinformático) a 80X de profundidad cuestan US $895/exoma*, el presupuesto sería de aprox./ $569 millones de pesos.(Ver cotización en: http://www.dnadtc.com/products.aspx)

Los pacientes de Tumaco, Guapi y Buenaventura serán tomados de las consultas remitidas a las ciudades capitales de los respectivos departamentos.

Dentro del grupo de tumores se analizaran 41 exomas de origen familiar. Este cálculo fue hecho con base en que el cáncer de mama familiar se da en un 15% de los casos, en consecuencia, se tomaran 41 exomas de ese tipo de cáncer. Esta proporción se extenderá para calcular el tamaño de la muestra de cáncer familiar para cada Dpto.

Argumentación bibliográfica: Dos artículos científicos recientes (Tamborero et al., 2013. *SCIENTIFIC REPORTS.* 3:2650 | DOI: 10.1038/srep02650; Lawrence M.S. et al., 2014. *Nature*. 505: 495) usaron aproximadamente 267 y 225 muestras de tumor mamario o exomas, respectivamente. Nuestro valor de 276 muestras de tumor se encuentra dentro de estos tamaños de muestra.

4.2.5 Características de los pacientes y número de exomas

Se plantea hacer un estudio epidemiológico descriptivo, observacional tipo prevalencia/transversal, con controles en pacientes del sexo femenino afectadas con Ca mamario familiar y Ca mamario esporádico según proporciones establecidas en la Tabla 5. En pacientes en etapas II y III (70-80% de los casos) del tipo celular intraductal y del tipo histológico infiltrante o invasivo (70-80% de los casos) y con rango de edad de 20 a 80 años. En este estudio se debe calcular el tamaño de muestra, con el fin de hacer asociaciones y extrapolar a la población en general. Para tal efecto, se analizarán 321 exomas de 276 pacientes y 45 individuos control. Adicionalmente el número de pacientes de cada región del SO se escogera de acuerdo a las proporciones establecidas en la Tabla 5.

Tabla 5. Número de pacientes y exomas a analizar

Cáncer de mama en el SO Colombiano				
Depts.	Número Mujeres	# de casos/año (2002-2006)	# Exomas o Pacientes	# Exomas Control
Nariño	408573	126	26	9
Cauca	325352	112	24	7
Valle	1128101	1039	220	26
Choco	119532	24	6	3
	1981558	1301	276	45

De cada paciente se tomará una muestra de tumor de mama; de los individuos saludables se tomaran 15 mL de sangre. Para los casos familiares que serán parte de este estudio (41 casos), se van a tomar muestras de tejidos, a) Tumor cuando el familiar del paciente este afectado por la misma patología, b) De tejido tumoral benigno, cuando el paciente tenga una patología mamaria benigna y sea necesario extraerla, c) De sangre (15 mL), cuando el familiar este sano y no tenga la patología tumoral maligna mamaria. Los resultados también serán comparados con las bases de datos públicas de exomas secuenciados en otras latitudes para el cáncer mamario. Finalmente, las variantes genéticas relacionadas serían las candidatas para el desarrollo de un potencial chip de diagnóstico y posterior “screening” poblacional en una fase III de factibilidad.

4.2.6 Toma de las muestras

Muestreo: Se tomaran 321 muestras de exomas de 276 pacientes afectados de cáncer o relacionados familiarmente a pacientes afectados de cáncer (ver Tabla 5).

Factores de Inclusión:

a) Estar afectado de cáncer mamario, en etapa II o III, y no haber aun recibido ni quimio ni radioterapia.
b) Familiar de un paciente afectado por cáncer, en las condiciones como se describe en a)
c) Familiar con una patología benigna mamaria.
d) Familiar sano
e) Individuo no familiar, sano, sin patología mamaria maligna o benigna (Control).

Factores de Exclusión:

a) Paciente afectado por otro cáncer, diferente al de cáncer de seno.
f) Familiar de un paciente afectado por otro tipo de cáncer.

A las personas candidatas para el estudio se les pedirá participar bajo la firma libre y voluntaria de un formato de Consentimiento Informado (CI), (se anexan los modelos de formatos). En el CI que para este estudio se adecuó, se hace especial énfasis en el objetivo de la firma de este CI que se usará, según cada caso, es decir, según los tipos de sujetos (pacientes, familiares y controles) que harán parte de la actual investigación.
Por lo anterior, para la toma de las muestras que se analizaran en este proyecto, se elaboraron cuidadosamente 5 modelos de CI que corresponden a cada caso particular del muestreo y que fueron aprobados por los Comités de Ética de las instituciones de salud que avalan y/o que acompañan esta propuesta:

-Consentimiento Informado 1: Para los pacientes con cáncer mamario que harán parte del estudio (se tomara muestra tumoral)

-Consentimiento Informado 2: Para los familiares afectados por la misma patología de los pacientes del estudio (se tomara muestra tumoral).

-Consentimiento Informado 3a y 3b:

3a. Para los familiares sanos de los pacientes del estudio que tienen una patología benigna de mama (se tomara muestra de tejido clasificado como benigno).

3b. Para los familiares sanos de los pacientes que no tienen ninguna patología benigna de mama (se tomara muestra de sangre).

-Consentimiento Informado 4: Para los sujetos controles que harán parte del estudio (se tomara muestra de sangre).

De cada paciente se tomaran los datos personales e historia clínica correspondiente los que se usaran para hacer análisis y correlaciones, de acuerdo a protocolos estándar y análisis estadísticos estándar.

4.3 Establecer un banco de muestras de ADN de pacientes y controles

4.3.1 Transporte, almacenamiento y banco de tumores

Las muestras de tumor serán transportadas en un termo con hielo seco. Posteriormente, una parte de la muestra será congeladas sin medio líquido a -20°C en un congelador (para replicas y banco de células).

4.3.2 Obtención de las muestras de tejido mamario

De la biopsia del tumor mamario extraída por el oncólogo tratante, se tomara un fragmento de aproximadamente 0.5 cm^3, para su posterior extracción de ADN genómico usando el kit QIAgen y secuenciación de exomas. Brevemente, el patólogo estandarizará el protocolo a seguir por los patólogos de la región involucrados en el estudio. Existen 3 métodos para tomar la muestra de biopsia de mama. 1) con aguja trucut (se obtiene poca muestra), 2) por cuadrantectomia y 3) por mastectomía. Se tomaran muestras por los métodos 2 y 3. Esto para dar prioridad a la muestra del paciente y su diagnóstico. Una vez el patólogo defina mediante una prueba anatomo-patológico la muestra para el diagnóstico y la muestra para el estudio de exómica esta última se pondrá en un tubo Corning de 50 mL en medio Dulbecco-MEM. Posteriormente, las muestras de cáncer para dx por parte de Patología, serán congeladas mediante un criostato (para corte, fijar y colorear), secciones para corte histológico serán revisadas por el patólogo, fijadas en formol para dar el diagnóstico definitivo. Las muestras para exómica que contengan más del 90% de células tumorales serán usadas para el estudio.

4.3.3 Obtención de las muestras de sangre: Protocolo para extracción de ADN exómico

Solicitud de análisis: Whole Exome Sequencing

Se extraerán 10 mL de sangre en un (purple-top) Tubos de EDTA (s). Esta muestra se manipula a temperatura ambiente en un recipiente aislado. No se debe calentar o congelar. La muestra debe procesarse dentro de 48 horas. Del ADN purificado tomar al menos 20µg de ADN purificado (concentración mínima de 50ng/µL; A260/A280 de ~ 1,7). Los ADN cuantificados se guardan a -20°C en un congelador, resuspendidos en buffer TE 1X.

4.3.4 Protocolo para extracción de ADN exómico

El proceso de extracción de ADN se hará usando un kit de QIAgen para muestras de tejido siguiendo el protocolo descrito por el fabricante. Los tumores mamarios se someterá inmediatamente a hidrólisis con proteinasa K para extracción de ADN. Similar procedimiento se seguirá con los controles de sangre. Los ADN serán resuspendidos en buffer TE 1X, cuantificados y guardados a -20°C en un congelador. El método WES estándar necesita una muestra de 1,0 µg para una profundidad de 100X.

4.4 Determinación de las secuencias de los exomas de las muestras obtenidas de las personas involucradas en el estudio

Debido a que actualmente no contamos con la tecnología necesaria para hacer secuenciaciones masivas de exomas en el país, las muestras de ADN, serán enviadas a Georgia Tech Institute (Atlanta-USA) previa autorización del Ministerio Salud y Protección Social y cumpliendo con todos los protocolos y normas reglamentarias establecidas para el envío, entre las cuales se encuentran las aprobaciones de los Comités de Ética que avalan y apoyan el actual proyecto de investigación.

El protocolo de secuenciación compromete los siguientes pasos, representados en la Figura 7:

- Fragmentación de las secuencias de ADN
- Preparación de las bibliotecas o librerias Illumina
- Enriquecimiento del exomas
- Generación de agrupamientos
- Secuenciación y "basecalling: Para la secuenciación se utilizará una plataforma NGS, HiSeq 2500 System**.**

Figura 7. Diagrama de flujo para el protocolo de enriquecimiento de exomas.

4.5. Análisis Bioinformático básico de los exomas utilizando la suite Atlas 2

El procedimiento bioinformático básico compromete dos grandes pasos, representados en la Figura 8: Alineamiento de lecturas (reads) de secuencias a partir de ficheros BAM y Detección de la variación.

El tamaño de los archivos BAM depende del cubrimiento (el número de veces que cada base es leída). La Tabla 6 a continuación da un ejemplo.

Table 6. Sobrevisión de los requerimientos de almacenamiento dependientes del cubrimiento o número de lecturas y longitud de cada lectura.

	Cubrimiento	No. of lecturas	Long lectura	Tamaño fichero BAM	Tamaño Avadis NGS
Exoma	80x	220,000,000	75	11.4 GB	14.2 GB

El análisis de regresión logística multifactorial (RLM) se da en el marco del análisis de las variantes. Estos es, los archivos BAM (Binary Alignment Map) de las SNPs y las InDels identificadas, son sometidos al mismo algoritmo para calcular la RLM y generar los archivos en formato VCF (Variant Call Format). Brevemente, se compila cada sitio de cada variante leída. Los datos compilados alimentan el modelo de RLM y las variantes que son de suficiente y alta calidad y pasan el filtro de la heurística, son luego genotipificadas y salen como archivos VCF. El siguiente diagrama de flujo muestra el orden secuencial del análisis

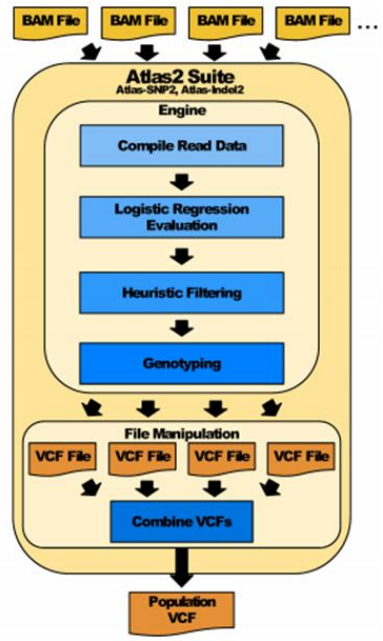


Figura 8. Diagrama bioinformático básico para secuenciación de exomas mediante la suite Atlas2 (Challis et al., 2012).

Para la RLM se tiene en cuenta el modelo de variable independiente (X) y variables dependientes e interacciones para los modelos SNP and INDEL junto con la estadística Wald Z de las variables y el valor p estimado mediante la función *glm* del medio ambiente R. Los valores Z y los valores p indican la significancia de las variables en el modelo y las variables más significativas tendrán un valor Z cuando se aleja de 0. Para información adicional consultar: Challis et al., 2012.

4.5.1 Análisis básico de exomas por Bioinformática utilizando el paquete SOAP (Opcional)

La secuenciación exómica generará miles de fragmentos cortos de ADN (de 100 a 500pb de longitud), los cuales corresponden a secuencias de ADN-codificante de los aproximadamente 23.000 genes que conforman el genoma humano (haploide) de cada uno de los individuos del estudio; lo que equivale a unos 189.000 exones o ~50 Mpb. Dentro de las estrategias de análisis se debe considerar la organización sistematizada de esta información bajo un genoma referencia (GRCh37 o 38), que permita dar certeza de la variabilidad genética de cada individuo, es decir para la identificación de variantes génicas y alélicas entre todos los individuos. La siguiente metodología describe cuatro (4) procedimientos bioinformáticos convencionales que serán aplicados dentro del presente proyecto, para los cuales se requiere la utilización de una plataforma de cómputo de alto nivel y ejecución, y personal especializado. Estos están integrados en el siguiente diagrama de flujo:

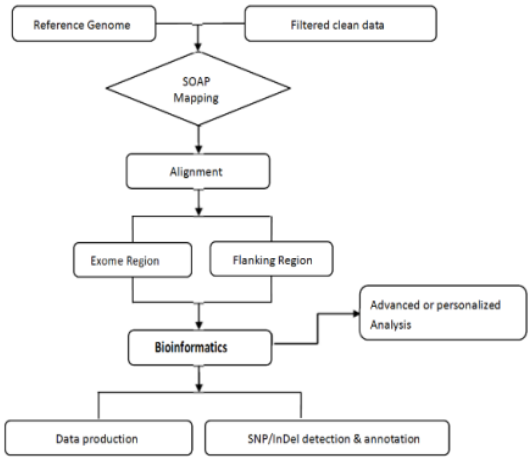


Figure 4: Pipeline for bioinformatics analysis.

4.5.2 Diseño experimental para la captura de exones que conforman el exoma

Se tomaran las coordenadas de los exones y de las regiones adyacentes a extraer (Exones + 5´UTR + 3´UTR + Primer intrón + región de control (2Kbp)) para cada uno de los genes del genoma humano, según coordenadas del genoma referencia GRCh37 o 38. Todas estas posiciones serán identificadas por Bioinformática y enviadas a la firma comercial o institución contratada (Georgia Tech o 23 and me, etc), para sintetizar las secuencias cebadoras para extraer los ~189.000 exones. Todas estas secuencias serán usadas para montar las plataformas para extraer el exoma de cada paciente a partir de las muestras de ADN genómico enviadas para tal fin.

4.5.3 Mapeo de secuencias cortas

El mapeo de secuencias cortas involucra el apareamiento de las secuencias propias con un genoma humano referencia, que para nuestro caso es el genoma depositado en la NCBI (versión actual: GRCh37). Dicho apareamiento puede realizarse mediante pathlines de alineamiento propio (Programación distribuida sobre algoritmo MAFFT), o mediante herramientas preexistentes como *Burrows–Wheeler Alignment Tool* (BWA) (Li y Durbin, 2009), o Maq (Li *et al*., 2008). Adicionalmente, se debe representar la densidad de información por región codificante del genoma, para esto se aplicaran estadísticos estándar, y herramientas gráficas de análisis como *Genome Analysis Toolkit* (GATK) (Aaron *et al*., 2010).

4.5.4 Identificación de variantes

La identificación de variantes implica un reto técnico, debido a que para este procedimiento se deben considerar y variaciones relevantes o no relevantes, dependiendo de la misma diversidad alélica de la especie, esto sumado a la comparación de las datos propios con los datos del proyecto 1000 genomas humanos, y otros que aparezcan durante el desarrollo de este proyecto. De esta forma, se realizan descripciones de las variantes genéticas a escala de la población humana. Este procedimiento se denomina: i*dentificación de variantes a partir de múltiples muestras*, para lo cual se requiere alinear y mapear las secuencias de forma jerárquica, en donde la jerarquía la define la descripción poblacional del individuo, este trabajo se realizará con herramientas propias y la utilización de herramientas como GATK-Modulo Unified Genotyper (Aaron *et al.,* 2010) y SAM tools (Li 2009).

4.5.5 Anotación de variantes

La anotación de variantes se refiere a las asociaciones de bases de datos funcionales (Ontology, KOG, KEGG, Wiki Pathways, etc) a cada variante. A partir de esta anotación es posible identificar posible impacto de la variante, en el contexto de la expresión génica, función proteica y metabolismo. Esta anotación se realizará mediante el desarrollo de pathlines de anotación propios, y la utilización de herramientas convencionales como ANNOVAR (Wang *et et al.,* 2010)

4.5.6 Descripción de susceptibilidad de cambio proteico por variantes

A partir de la anotación, se identificarán los procesos biológicos críticos afectados por las variantes. Estos serán analizados a partir de la descripción estructural y funcional de las proteínas normales y mutantes relacionada a las variantes, Para esto se crearán herramientas propias y se utilizarán herramientas convencionales como PolyPhen-2 (Adzhubei *et al*., 2010) y SIFT (Kumar *et al*., 2009).

4.6.7 Simulación de corrida de exomas para evaluar los requerimientos computacionales

La secuenciación exómica se realiza a partir de una librería construida por fragmentación del genoma completo a estudiar. A continuación, se capturan los fragmentos del genoma que corresponden a los exones mediante un array con sondas específicas para dichas regiones. Una vez capturados todos los exones (o, hoy por hoy, la mayor parte de ellos, como explicamos más adelante) se procede a secuenciarlos; finalmente, por una serie de procedimientos bioinformáticos, estas secuencias se vuelven a mapear sobre el mapa del genoma de referencia y se pueden apreciar los cambios (mutaciones) existentes. La ventaja que ofrece este abordaje (frente al más utilizado hasta hace muy poco, basado en arrays de SNPs) es que permite capturar mucha más información, pues no sólo identifica los SNPs asociados a la enfermedad, sino también las propias mutaciones causantes o directamente asociadas a ella.

4.5.8 Remapeado de Exoma por métodos bioinformáticos (SOAP: Short Oligonucleotide Analysis Package)

SOAP es la evolución de una única herramienta de alineación a un conjunto de herramientas que ofrece una solución completa a la nueva generación de datos de secuenciadores. Con SOAP se remapea por completo los datos obtenidos de secuenciadores de última generación y ultra alta eficiencia.

En la actualidad, se trata de una nueva herramienta de alineación (SOAP aligner/soap2), constructor secuencia consenso y re-secuenciación (SOAP snp), un buscador de indel (SOAP indel), un escáner de variación estructural (SOAP sv) y un ensamblador de novo de lecturas cortas (SOAP de novo). Y una herramienta de alineación acelerado por GPU (SOAP3/GPU) que se encuentra en implementación.

Como el volumen de información a tratar es elevado los requerimientos computacionales también lo son ya que dependen directamente de la cantidad de datos que se quieran procesar.

Para la evaluación de un solo exoma en CPU.

Requerimientos mínimos de sistema:

- **Hardware:**

a) 64-bit x86-64 CPUs with SSE instructions.
b) 8 GB main memory (for a genome as large as human's).
c) 8 GB hard disk (for a genome as large as human's).

•**Software:**
a) 64-bit Linux system (kernel >=2.6).

Evaluación de rendimiento:

SOA Paligner necesita cerca de 2 horas para dar formato a la secuencia de referencia y construir tablas de indexación. El uso de la RAM depende del tamaño total de la secuencia de referencia. Para la referencia del genoma humano, ocupará 7 GB RAM, Tabla 6.

Tabla 6. Rendimiento de alinear 1millon pb leyendo un solo extremo (longitud lectura 35pb) o 1 millón de pares de bases en genoma humano como referencia.

	Time (sec)Single-end reads	Time (sec)Paired-end reads	RAM (GB)
SOAPaligher(soap2)	120	505	6.8
Soap	1700+	5743	13.4

Estos serían los datos para efectuar el mapeado completo en CPU.

Si un exoma humano promedio tiene 50 Mb de longitud, y usamos una RAM de 6.8 GB, tardaría 420 min (7 horas por exoma). Para los 300 exomas serían 87 días de trabajo ininterrumpido. Si aumentamos la RAM a 200 GB (lo que deseamos adquirir) o 1500 GB (lo que hay en BIOS-Manizales), el trabajo se haría aproximadamente en algunas decenas a pocos días de procesamiento, respectivamente. Si a esto le sumamos los análisis comparativos con los mas de 25.000 bases de datos de genomas de cáncer de otros países, bien podremos hablar de varios meses de proceso *in silico*.

4.5.9 Comparación de los genes y las variantes genéticas halladas con las bases de datos de cáncer disponibles públicamente

Las secuencias de exomas obtenidas serán comparadas mediante análisis BLAST y otros algoritmos con las secuencias de genomas de más de 25.000 tumores que se encuentran depositadas en las bases de datos públicas de países desarrollados (Ledford, 2010), a fin de evidenciar la frecuencia o novedad de los genes y variantes encontradas.

4.6 Análisis estadístico

Se usaran técnicas estadísticas estándar y especializadas para anotar y asignar las probabilidades. Esto es: 1) Una clasificación de mutaciones potencialmente deletéreas. 2) Se hará una identificación de mutaciones "conductoras" de cáncer (cancer driver mutations). Opcionalmente, una identificación de mutaciones que puedan ser blanco de medicamentos (druggable mutations) podrá ser adicionada.

4.6.1 Una clasificación de mutaciones potencialmente deletéreas

Las secuencias serán evaluadas computacionalmente. La existencia de variantes será evaluada usando la dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) y si no está presente, será directamente evaluada en el DNA normal del mismo individuo. El ADN normal (tejido sano o linfocito) confirmaría la naturaleza somática de la mutación y eliminaría los artefactos de secuenciación. El análisis estadístico involucra calcular la desviación de la razón de mutaciones sinónimas: no sinónimas a partir de lo esperado por azar y será usado para indicar la presencia de selección sobre las mutaciones no sinonímicas. La diferencia entre el número observado y esperado de las mutaciones no-sinonímicas, a través de todas los tipos de mutaciones está dada por:

$$Obs - Exp = \sum_k \left(n_k - t_k \frac{N_k}{T_k} \right)$$

Para evaluar la significancia de esta razón, se hará una prueba Monte Carlo, la cual es aplicada al set completo y a los sub-sets de mutaciones.

4.6.2 Identificación de mutaciones "conductoras" de cáncer

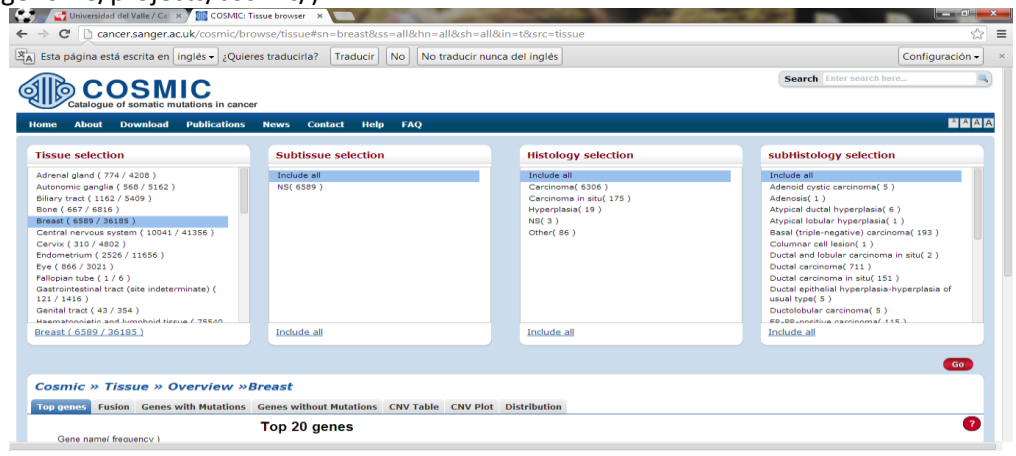
El número de mutaciones conductoras se determina mediante la cuantificación de las presiones de selección y el número de mutaciones conductoras implicadas en el desarrollo del cáncer. La presión de selección es definida como ϕ. Esta representa el incremento relativo en probabilidad de observar una mutación no sinónima debida a selección positiva por cáncer. Mediante condicionar el número de conteos totales de las mutaciones t_k, la distribución de mutaciones sinónimas y no sinónimas obedece a una distribución binomial:

$$\Pr(\{s_k, n_k\}_k) = \prod_k \frac{t_k!}{s_k! n_k!} \frac{(S_k)^{s_k} (\phi N_k)^{n_k}}{(S_k + \phi N_k)^{t_k}}$$

Donde, la presión de selección ϕ es estimada vía probabilidad máxima. Para calcular la presiones de selección específica de los genes, esta técnica considera los conteos silentes a través del set completo de todos los genes (este asume que la razón de mutaciones subyacente es constante a través de todos los genes. En este punto nos concentraremos en 145 a 150 genes reportados como conductores de cáncer en la literatura, más otros (presumibles) que podrían estar involucrados. Finalmente, las bases de datos de rutas metabólicas mediante Reactome, Panther e INOH nos podría ayudar a identificar la presencia de rutas metabólicas mutadas (doi: 10.1038/nature05610).

4.6.3. Comparación de los genes hallados con bases de datos públicas

Los genes y las mutaciones halladas se compararan con los genes depositados en las base de dados del Catálogo de Mutaciones Somáticas en Cáncer, COSMIC (http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/).



4.7 Análisis Multifractal y análisis de discriminación

Uno de los problemas críticos por resolver en el análisis masivo del genoma humano y su expresión es el impacto de este conocimiento en la salud humana (personalizada). A la fecha hay una gran controversia acerca de la utilidad práctica de la secuencia del GH para el médico (http://www.bbc.co.uk/mundo/noticias/2013/12/131211_salud_genoma_conocer_enfermedades_ap.shtml). El problema surge por la falta de una interpretación real, adecuada y predictiva del comportamiento saludable o patológico del GH. El siguiente enfoque es un esfuerzo complementario al análisis WES, abordado desde la geometría fractal, a fin de contribuir a interpretar –de la mejor manera posible- la secuencia del GH. Para ello se ha propuesto un enfoque novedoso llamado: *Bioinformática multifractal* (Moreno, 2014), el cual se aplicará en el presente trabajo.

Brevemente, un fractal es una figura geométrica fragmentada, cuyas partes son una copia a escala aproximada de toda la figura, es decir, la figura posee auto-similitud. La dimensión fractal D de la figura es básicamente la regla de la escala que la figura obedece, En general, su distribución de frecuencias obedece a una ley de potencia. Un multifractal es un enfoque donde varios fractales coexisten, es decir, el multifractal se puede describir mediante múltiples leyes de potencia. En este proyecto, el análisis multifractal se empleará a fin de discriminar las variantes de los pacientes con Ca mamario de las variantes presentes en los controles, de manera similar como este análisis permite discriminar, pacientes con falla cardiaca, de personas saludables (Ivanov *et al.*, 1999). Los análisis de genómica multifractal (Moreno *et al*., 2011) se harán con el software BioCaos (Moreno *et al.*, 2006).

4.8 Desarrollo de los planos para la construcción de un chip para el diagnóstico del cáncer mamario.

33 genes han sido estar implicados con el cáncer mamario (Vogelstein et al., 2013). Sin embargo, el número de estos sigue en aumento. Hoy se estima que unos 140 a 150 genes conductores podrían estar asociados en menor y mayor grado con la génesis de todos los tipos de cáncer. Nosotros esperamos que el análisis de exoma revele algunos genes nuevos relacionados con el ca mamario. Para elaborar un potencial chip, nosotros proponemos 1) hacer un e**studio de las tecnologías de Micro-arreglos** y análisis detallado del análisis estadístico estándar. 2) **Efectuar un estudio de los factores de error en el procesamiento de resultados de los microarreglos.** Estudio de los distintos factores que contribuyen a aumentar el error en los resultados de los experimentos con micro-arreglos específicamente la geometría o distribución de los genes de prueba en el arreglo. 3) Un e**studio de la geometría fractal y su aplicación en los microarreglos:** Estudio de la geometría fractal y su aplicación en el diseño de microarreglos para disminuir errores en los resultados obtenidos al considerar diferentes distribuciones con auto-similitud de los genes de prueba. **4) Un estudio, selección y apropiación de las herramientas de simulación de microarreglos.** Estudio de distintas herramientas de simulación de microarreglos, selección de la herramienta más apropiada para los requerimientos del trabajo y apropiación del manejo de la herramienta para llevar a cabo las simulaciones requeridas. Y **5. Una simulación de los microarreglos** con diferentes configuraciones geométricas y análisis de datos obtenidos. Todo esto con el fin de disminuir la tasa de error de falsos positivos que se presentan en el análisis de microarreglos. Adicional a esto, se diseñaran estrategias para la deposición de las sondas sobre el chip y cuantificación y análisis de las emisiones fluorescentes mediante análisis de “box-counting” y multifractal.

4.9 Funciones de las instituciones y sus roles

Institución	Role	Actividades/Funciones	Organización
Universidad del Valle: EISC EIEE EIQ ECN (Biología) ECB (Microbiología)	Ejecutor	Contactar los oncólogos y patólogos, contactar los pacientes, encuestas, consentimiento informado, codificación de las muestras, toma de muestras biológicas, separación tejido tumoral-tejido sano, transporte de muestras, procesamiento (extracción de ADN) de muestras, almacenamiento de muestras biológicas, envío de muestras de ADN al exterior. Escritura de scripts. Procesamientos pilotos de programas de computación. Elaboración de bases de datos, elaboración de un sitio web. Análisis estadístico WES. Análisis multifractal. Implementación de los planos de un chip potencial. Escritura de artículos y documentos.	Dirección, coordinación y co-investigadores
Universidad del Cauca: Dpto. de Biología (Grupo BIMAC)	Colaborador	Contactar los oncólogos y patólogos, contactar los pacientes, encuestas, consentimiento informado, codificación de las muestras, toma de muestras biológicas, transporte de muestras, procesamiento (extracción de ADN) de muestras, almacenamiento de muestras, envío de muestras a la Universidad del Valle y Análisis exómicos que contribuyan al desarrollo del diseño del prototipo. La Universidad del Cauca participara de manera directa y continua durante el desarrollo de proyecto para obtener los resultados esperados, además los investigadores participaran en escritura de artículos y documentos, por tanto serán coautores de estos.	Co-investigadores
CBBC - Manizales	Colaborador	Procesamiento de datos. Análisis masivo de exomas y con otras bases de datos públicas. Colaboración en la escritura de artículos y documentos.	Co-investigadores
Georgia Tech Dpt. of Biology. Dpt. of Genetics	Asesor internacional	Desarrollo de las plataformas para WES. Secuenciación de exomas. Transferencia de tecnologías (envío del paquete SOAP optimizados por ellos) para análisis de exomas. Colaboración en la escritura de artículos y documentos.	Co-investigadores
Hospital Universitario	Colaborador	Proveer pacientes y muestras de tejidos	Co-investigadores

del Valle			
Clínica Imbanaco	Colaborador	Proveer pacientes y muestras de tejidos	Co-investigadores
Secretaria Departamental de Salud del Valle	Colaborador	Información en salud del cáncer mamario	Co-investigadores
Hospital San Rafael, Zarzal	Colaborador	Información en salud del cáncer mamario. Proveer pacientes	Co-investigadores
Secretaria de Salud del Valle del Cauca	Colaborador	Información en salud del cáncer mamario y epidemiología	Co-investigadores
Secretaría de Salud de Bogotá	Colaborador	Información en salud del cáncer mamario y epidemiología	Co-investigadores
Pontificia Universidad Javeriana	Colaborador	Asesoría en Bioinformática y computación	Co-investigadores
Hospital San José, Popayán	Colaborador	Proveer pacientes y muestras de tejidos	Co-investigadores
Hospital Dptal. San Frco. de Asis	Colaborador	Proveer pacientes y muestras de tejidos	Co-investigadores
Hospital Dptal. San Frco. de Asis	Colaborador	Proveer pacientes y muestras de tejidos	Co-investigadores
Hospital Universitario Dptal. de Nariño	Colaborador	Proveer pacientes y muestras de tejidos	Co-investigadores

5. IMPACTOS Y BENEFICIOS

1. Impacto regional

- Conocer por primera vez una muestra del estado exómico de una muestra de nuestras gentes y su relación con el cáncer de seno.
- Proveer a la región del SOC del país con una plataforma genómica y computacional para el apoyo del análisis médico del cáncer. Un ecosistema computacional de alto rendimiento (hardware + software) que sirva como herramienta para el estudio y apoyo al diagnóstico médico del cáncer de seno.
- Generar nuevas alternativas para la prevención de cáncer que permitirá el desarrollo humano contemplando, en el *disfrutar de una vida prolongada.*

2. Impacto económico

- Desarrollar una metodología preventiva que contribuya a reducir los costos en el sistema de salud y mejorar la calidad del capital humano productivo de la región.
- Busca generar potenciales productos (chips) útiles en la prevención de Cáncer que mitiguen los altos costos de tratamientos.

3. Impacto social

- Esta iniciativa propende y está interesada en producir tecnologías y conocimientos innovadores o creativos, a través de la conformación de una plataforma para la producción de conocimiento integrada por diferentes instituciones nacionales e internacionales, que permitan a futuro aportar nuevos empleos y nuevas soluciones con aplicación en la salud. Además de la formación de profesionales en tecnología de punta y el desarrollo empleos y nuevas fuentes de ingreso para la región que impactaran en la calidad de vida de sus habitantes.
- Contribuir a mejorar los niveles de salud de la región
- Recopilar información genómica útil como base para diseñar futuros estudios, más extensos.

4. Impacto institucional

- Las instituciones involucradas serán pioneras y líderes en desarrollar estudios masivos en ciencias ómicas y salud, lo cual dará visibilidad nacional e internacional a nuestra Plataforma.
- Formación de recursos humanos al más alto nivel ciencias ómicas y salud.

En términos de beneficios,

1. Beneficios sociales:

-Generación de una tecnología con potencial de hacer un dx temprano y presintomático (susceptibilidad genética) de cáncer de mama.

-Reducción de la tasa de morbilidad y mortalidad de algunas de las personas beneficiadas por el proyecto.

-Contribuir a incrementar el ingreso *per capita* de una familia, donde las personas potencialmente blanco de la enfermedad son contribuyentes.

-Medicina personalizada que evite el empleo excesivo, deficiente o equivocado de los tratamientos que contribuyen a la mortalidad.

-Utilización de nuevos genes y "perfil" genéticos personalizados para cada paciente.

-Estos estudios se pueden extender fácilmente (por nosotros) a otros tipos de cánceres y enfermedades de carácter genómico.

-El macroproyecto apoya a las propuestas de los planes de desarrollo nacionales, departamentales y municipales.

-Que las aplicaciones lleguen rápidamente al consultorio médico.

2. Beneficios Económicos: Contribuir a disminuir los costos de diagnóstico y tratamiento.

3. Beneficios Ambientales: Contribuir a mejorar la calidad de vida de las personas afectadas.

6. Resultados y Productos esperados

Objetivo	Resultado o Producto
COMPONENTE 1: 1.1 Identificación de las personas 1.2 Obtención de las muestras de sangre de controles y tumores de cada paciente 1.3 Banco de tumores mamarios y Banco de ADN 1.4 Realización de campañas de sensibilización y concientización sobre la participación en el estudio en ciencias Ómicas	-Registro de personas en las entidades de Salud Informes, registro, historia clínica de personas y diagnósticos. -Consentimientos informados -Muestras biológicas congeladas -ADN genómico cuantificado -Permiso del Min de Salud para sacar las muestras de ADN del país

COMPONENTE 2: 2.1 Implementación de protocolos y secuenciación de los exomas 2.2 Adquisición de una estructura computacional escalable para análisis bioinformático.	- Plataforma computacional Ilumina para la extracción y secuencia de los exomas. -Secuencias de los exomas -Establecimiento de una infraestructura computacional funcional
COMPONENTE 3: 3.1 Analizar e interpretar con métodos bioinformáticos, genéticos, biomédicos y enfoque multifractal las variantes de secuencia en exomas humanos. 3.2 Elaboración de un software de análisis que permita el estudio automatizado y personalizado de exomas 3.3 Elaboración de un software integrado para análisis multifractal de exomas y genomas humanos 3.4 Elaboración de un portal de información para la población experta y población en general, del Sur Occidente Colombiano y del mundo, relacionado con las enfermedades 3.5 Enfoque multifractal	- Variantes genéticas y genes conductores. - Un software de análisis que permita el estudio automatizado y personalizado de regiones de genoma (instalado en la plataforma Galaxy). - Portal de internet FTP de bases de datos y HTTP de exomas secuenciados para la población experta (médica) y población en general relacionado con el cáncer mamario. - Un software integrado para el análisis multifractal de regiones de genomas humanos
COMPONENTE 4: 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.	- Genome Browser de la Plataforma en Ciencias ómicas y salud del Cáncer mamario en Colombia - 5 artículos científicos publicados
COMPONENTE 5 5.1 Elaboración de un diseño computacional 5.2 Simulación y modelamiento de los procesos computacionales 5.3 Verificación y validación del diseño computacional	- Diagrama del diseño de un prototipo de diagnóstico en chip

NOTA: Los 7 objetivos específicos del proyecto fueron resumidos en esta tabla de Resultados esperados en 5, a fin de facilitar la introducción de los datos en la MGA. De lo contrario, hubiéramos tenido que hacer una nueva MGA y no hay tiempo para esto.

7. TRAYECTORIA DE LOS GRUPOS DE INVESTIGACIÓN QUE CONFORMARAN LA PLATAFORMA

- Grupo de Bioinformática (GBioinfo):

El grupo pertenece a la EISC: *http://bioinformatica.univalle.edu.co*, por lo tanto, nuestros enfoques son principalmente orientados a informática y computación. Nosotros estamos fascinados por preguntas microbiológicas-moleculares y sus complejidades y buscamos contribuir a su análisis y solución.

Nuestros tópicos principales incluyen predicción de genes y proteínas, donde nosotros aplicamos métodos de inteligencia artificial con base en técnicas de aprendizaje de máquinas, como también técnicas de minería. Nosotros también trabajamos con la teoría del caos y el análisis multifractal para el análisis de imágenes médicas y genómicas (Wikipedia: Multifractal system, 2012) a fin de

predecir el comportamiento de las funciones y estructuras codificadas por los genomas. También trabajamos en analizar el medio ambiente de los Andes Colombianos, mediante metagenómica a fin de descubrir nuevos productos (bioprospección). Generalmente, nuestros resultados incluyen herramientas en software las cuales permiten a los usuarios aplicar nuestros enfoques a sus objetos de investigación.

Nuestro grupo es interdisciplinario, conformado por profesores y estudiantes de pre y postgrado de diferentes disciplinas de las ciencias de la computación y la biología. Esto nos facilita abordar problemas complejos usando tanto: el estado del arte del conocimiento biológico, como las metodologías computacionales avanzadas.

- Irene Tischer (Líder del grupo, Profesora de la Facultad de Ingeniería)
- Pedro A. Moreno (Jefe de investigación, Profesor de la Facultad de Ingeniería)
- Oscar Bedoya (Investigador, Profesor de la Facultad de Ingeniería)
- Margot Cuarán (Investigador, Estudiante de doctorado)
- Luis Garreta (Investigador, Estudiante de doctorado)
- Diego Mejía (Investigador, Estudiante de doctorado)
- Nilson Mossos (Investigador, Estudiante de doctorado)
- Fabián Tobar (Investigador, Estudiante de doctorado)
- Andrés Becerra (Investigador, Estudiante de doctorado)
- Oscar Restrepo (Investigador, Estudiante de maestría)

Campos de investigación:

Nuestro grupo trabaja en varios campos de investigación listados abajo. Nuestra meta en todos es: proponer análisis computacionales y de predicción novedosos y mejorar los existentes en cada campo de investigación; contribuir a la educación en bioinformática a nivel de pregrado y posgrado y soportar con enfoques informáticos fuertes y bioinformática, la investigación en biología computacional: Modelamiento y predicción de genes y genomas, modelamiento de la función y la estructura de la proteína, metagenómica y desarrollo de software bioinformático.

Proyectos de investigación: El GBioinfo ha participado en varios proyectos de investigación: 1) Análisis multifractal del genoma humano, proyecto Colciencias #1103-12-16765 (2006-2010). 2) Centro de Excelencia GeBiX para el análisis genómico y bioinformático de ambientes extremos, proyecto Colciencias # 6570-392-19990 (2008-2014) y 3) proyectos de convocatorias internas de la Universidad del Valle con el GBNE.

- A metagenomics bioinformatics platform for characterization and exploitation of genetic resources in Colombian extreme environments. Centro de Investigación de Excelencia en Genómica y Bioinformática, 2008-2013.

- A graphic environment for eukariote gene prediction based on a highly configurable software tool (Genezilla). Bioinformatics and Biocomputation, Universidad del Valle (2011)

- Optimization of eukariote gene prediction, applying decision trees and bayesian nets. Bioinformatics and Biocomputation, Universidad del Valle (2011)

- Genomic cartography of the integration process of provirus HTLV-1. Laboratorio de Biología Molecular y Patogénesis; Bioinformatics and Biocomputation, Universidad del Valle (2009)

- Molecular and genome analysis of sequences and complete genomes. BIMAC-Universidad del Cauca; Bioinformatics and Biocomputation, Universidad del Valle (2009)

Cursos ofrecidos: bioinformatics I, restricted molecular dynamics, machine learning in bioinformatics, stochastic models in bioinformatics, introduction to bioinformatics y bioinformatics tools

Recursos computacionales:
El GBioinfo cuenta con dos pequeños laboratorios de cómputo con 16 work-stations y cuatro pequeños servidores. No obstante, la EISC proyecta adquirir servidores con una mayor capacidad de almacenamiento y de procesamiento.

- Grupo de Biología Molecular y Patogenicidad (GBMP): Dirigido por el profesor y Director científico Felipe García-Vallejo. Los principales logros de conocimientos científicos y desarrollo tecnológico han sido:
1. El análisis del genoma de virus de importancia para la salud humana en nuestro medio. La aplicación de las técnicas de ingeniería genética, permitió rápidamente la clonación del genoma de ciertos virus humanos principalmente los respiratorios tales como los adenovirus, el Virus Respiratorio Sincicial (RSV), además del HTLV-1, VIH-1 y del VPH, cuyas infecciones representan problemas de salud pública importantes en nuestro País.
2. El estudio de la epidemiología molecular de virus humanos de importancia en salud pública en nuestro país; en este sentido hemos analizado la circulación de subtipos de los adenovirus respiratorios causantes de brotes epidémicos de infección respiratoria en varias regiones de Colombia. En otra línea de investigación se inició el estudio de comparación de secuencias de nucleótidos de porciones de genes de las cepas de HTLV-1 que están circulando en varias poblaciones de la costa pacífica Colombiana que reveló aspectos importantes de la genética y evolución de este virus.
3. Desarrollo y aplicaciones bioinformáticas para analizar la dinámica de la integración de los retrovirus en el genoma de las células infectadas. Mediante la utilización de herramientas bioinformáticas y la construcción de bases de datos producto de una extensiva minería de datos, se han analizado las características estructurales y funcionales de extensas regiones del genoma humano donde ocurre la integración del DNA viral.
4. Exploración de productos naturales derivados de plantas colombianas para su utilización como antirretrovirales. Se ha demostrado que extractos de café ricos en DCQAs, presentan una potente actividad inhibitoria *in vitro* de la integrasa del HTLV-I. Esto constituye un logro importante pues abre una nueva frontera de aplicación terapéutica mediante la inhibición de esta enzima retroviral.
5. Desarrollos biotecnológicos en la implementación de producción de proteínas recombinantes para su utilización en estuches diagnóstico y en la producción de bioindicadores. Se han clonado y producido tanto en sistemas procarióticos como en eucarióticos, las proteínas de las envolturas virales del HTLV-I y del VIH-1; además se ha logrado la producción en el sistema de baculovirus, de una Integrasa recombinante funcional del HTLV-I; con base en esta enzima recombinante se diseñaran y desarrollaran métodos de bioindicación para el monitoreo de la infección por retrovirus humanos.
6. La exploración de la biodiversidad de organismos mediante enfoques moleculares. Se han incluido estudios sobre especies de ranas venenosas del Choco del Género Dendrobates. La exploración molecular de la diversidad genética y estructura poblacional de especie del delfín rosado del amazonas, *Inia geoffrensis* y de poblaciones de *Anopheles darlingi* en nuestro País.

Todo esto validado por más de 130 publicaciones nacionales e internacionales, 12 libros y documentos científicos y 10 estudiantes de doctorado graduados.

Investigadores Titulares

Mercedes Salcedo. MSc. PhD. Ciencias Biomédicas. Universidad del Valle.
Profesora Asociada. Escuela de Bacteriología y Laboratorio Clínico. Facultad de Salud. Universidad del Valle.

Adalberto Sánchez. PhD. Texas A&M University USA. Profesor Asociado. Departamento de Ciencias Fisiológicas. Escuela de Ciencias Básicas Facultad de Salud Universidad del Valle.
Edwin Carrascal. MD. Patólogo. Profesor Titular. Departamento de Patología. Facultad de Salud. Universidad del Valle.
Martha C. Domínguez. MSc. Doctora en Ciencias Biomédicas. Universidad del Valle. Investigadora Asociada.
Julio Cesar Montoya. MSc. Doctor en Ciencias Biomédicas. Profesor. Facultad de Salud. Universidad del Valle. Profesor de la Universidad Autónoma de Occidente.
José Maria Satizabal. MD. MSc. Profesor Titular. Departamento de Ciencias Fisiológicas. Facultad de Salud. Universidad del Valle. (Candidato a Doctor en Ciencias Biomédicas).
Oscar Tamayo. MSc. Universidad del Valle. Profesional Facultad de Salud Universidad del Valle. Catedrático Universidad Santiago de Cali. Cali.
Yesid Cuesta Astroz. BS. Universidad del Valle. Joven Investigador. Facultad de Salud. Universidad del Valle. Valle
Elizabeth Londoño. Estudiante del Programa de Maestría en Ciencias Biomédicas. Facultad de Salud. Universidad del Valle. Valle
Juliana Soto Girón. BSc. Genética .Universidad del Valle
Veronika Ceballos. BSc. Genética. Universidad del Vallle
Dianora Fajardo. Estudiante del programa de Biología. Facultad de Ciencias
Ángela Viviana Peña González. BS. Joven Investigadora de Colciencias. Universidad del Valle

-**Grupo de Genética Molecular Humana (GHMH)**
Director: Dr. Guillermo Barreto Rodríguez
Creado: Enero de 1999.
Código Colciencias: COL0005412.
Estado: Categoría Reconocido.
Clasificado en: Convocatoria Año 2012.
Integrantes: 4 investigadores (3 PhD y 1 MSc), 4 estudiantes de doctorado, 8 estudiantes de Maestría, 5 estudiantes de pregrado, 4 jóvenes investigadores de Colciencias y una Técnica Profesional. Total: 26.

CONTEXTUALIZACIÓN DE LA ACTIVIDAD DE INVESTIGACIÓN DEL GRUPO.
Los grandes avances de la biología molecular en los últimos años y las perspectivas surgidas con el proyecto del genoma humano generaron la urgente necesidad de formar recursos humanos al más alto nivel a objeto de aplicar este conocimiento y estas tecnologías para dilucidar de la estructura genética (patológica y normal) de las poblaciones humanas y animales en general. Teniendo en cuenta la escasez de grupos de investigación en Colombia enfocados al estudio de esta innovadora temática nace en 1999, en la Sección de Genética del Depto. de Biología de la U. del Valle el Grupo de Genética Molecular Humana con la finalidad de dar respuesta a este desafío. El conocimiento de la Biodiversidad a nivel molecular de las poblaciones humanas colombianas y particularmente del Sur-Occidente colombiano se encuadran dentro de los intereses investigativos de la Universidad del Valle, el Depto. del Valle y de Colombia.

MISIÓN: Realizar investigación científica de alta calidad en genética humana.
VISIÓN: Ser reconocido como un grupo de investigación líder en Colombia considerando el impacto del conocimiento generado y la calidad del recurso humano formado.
OBJETIVOS:

- Generar nuevo conocimiento en el área de la genética humana mediante la aplicación de la tecnología del ADN.
- - Caracterizar la diversidad patológica y normal a nivel de ADN en las poblaciones humanas.
- Formar recursos humanos al más alto nivel (doctorado).
- Generar y hacer transferencia de tecnologías encaminadas al estudio del genoma humano.
- Realizar proyectos de investigación en colaboración con grupos nacionales y del exterior.

LÍNEAS DE INVESTIGACIÓN:
1. Caracterización de genes comprometidos con patologías hereditarias.
2. Estudio de la variabilidad genética a nivel de ADN en poblaciones humanas.
3. Caracterización de la diversidad genética animal.

SECTORES DE APLICACIÓN DEL CONOCIMIENTO GENERADO.
1. Cuidado a la salud de las poblaciones humanas.
2. Cuidado a la salud de las personas.
3. Desarrollo de productos tecnológicos para la salud humana.

FORMACIÓN RECURSOS HUMANOS
- **Dirección de tesis de doctorado terminadas**.

Estudiante: Laura Cifuentes. Título tesis "Detección de mutaciones en los genes BRCA1 y BRCA2 en pacientes con cáncer de mama familiar del Sur – Occidente cololombiano". Grado 2011.

Estudiante: Fernando Rondón González. Título de la tesis: "Estudio de la variabilidad genética en poblaciones humanas del centro y suroccidente colombiano mediante el uso de marcadores moleculares". Grado: 2009.

Estudiante: Luz Angela Alvarez. Título tesis: "Diversidad genética del ganado Hartón del Valle y sus relaciones con Holstein y Brahman, mediante el uso de marcadores moleculares". Grado: 2008.

- **Grupo de investigación de Percepción y Sistemas Inteligentes (GPSI)** apoya el anteproyecto: "Plataforma en ciencia ómicas y salud del cáncer mamario del SO" que conduzca a la creación de un centro de investigación especializado desde su campo del conocimiento y experiencia en la formulación de proyectos de instrumentación, comunicaciones industriales, procesamiento de señales e imágenes, inteligencia computacional y robótica. En el campo de la bioingeniería el grupo ha realizado en asocio con médicos especializados desarrollos sobre procesamiento de señales electromiográficas, electroencefalográficas y electrocardiográficas para prediagnóstico; así como en imágenes del espectro visible, Rayos Xs e Infrarrojo para detección de lesiones en la pie, en los ojos, fracturas óseas y cáncer en el tórax y seno. El grupo creado en el 2000 está integrado por nueve profesores con formación doctoral y de maestría y en la actualidad cuanta con cinco estudiantes de doctorado y doce de maestría. El Grupo está en capacidad de apoyar proyectos de Infraestructura de Medición Avanzada, Ensayos No-Destructivos, aplicación de técnicas de inteligencia computacional en el modelado de sistemas y equipos con base en el conocimiento alojado en los datos de medición y de operación.

-Grupo de Bionanoelectrónica (GBNE).
Director: Jaime Velasco Medina.

Objetivo Principal:

Diseño computacional de bionanosistemas (bionanosensores, bionanodispositivos y bionanomáquinas) basados en nanotubos de carbono, grafeno, polímeros o proteinas para diagnóstico y tratamiento médico usando biofármacos y vacunas moleculares.

Objetivo Secundarios:

a) Fabricación de bionanosensores (bioNEMS), microsensores (bioMEMS) y bionanodispositivos (biochips-microarrays) para el diagnóstico y tratamiento médico usando las facilidades de los entes de cooperación internacional.

b) Diseño del hardware integrado para el procesamiento de la información generada por los bionanosensores, microsensores y bionanodispositivos.

c) Diseño de una arquitectura hardware para el secuenciamiento de ADN basado en el análisis multifractal.

Experticia

El Grupo de Investigación en Bionanoelectrónica de la Universidad del Valle, formado en 2002, clasificado como categoría D en Colciencias, tiene como principales líneas de investigación el diseño computacional de bionanosistemas (bionanosensores, bionanodispositivos y bionanomáquinas), y el diseño de sistemas integrados complejos basados en tecnologías reprogramables como FPGAs, FPAAs y PSoCs. El grupo tiene una amplia experiencia de investigación y desarrollo tecnológico, demostrada por más de 160 artículos publicados en los últimos 10 años en revistas y eventos nacionales e internacionales. Actualmente, el grupo cuenta con 5 profesores (2 profesores tiempo completo y 3 profesores hora cátedra), 4 estudiantes de doctorado con beca de Colciencias, 4 estudiantes de maestría y 15 estudiantes de pregrado, que desarrollan actividades de investigación.

El grupo tiene una comprobada experticia en el diseño computacional de bionanosistemas usando métodos como elementos finitos, dinámica molecular, DFT, dinámica cuántica, en diferentes entornos de simulación como ANSYS, Gaussian, Jaguar, LAMMPS, SIESTA, etc, para el análisis de las propiedades mecánicas y electrónicas. En esta línea de investigación se han diseñado bionanosensores basados en CNT-DNA, heteroestructuras CNT-BN, nanoantenas de grafeno, nanomáquinas de CNT-Lonsdaleita, etc. En este proyecto se pretende contribuir con el diseño bionanosensores (BioNEMS) y microsensores (BioMEMS) para la detección de enfermedades prevalentes en el Valle del Cauca con una alta tasa de mortalidad y un alto costo para el sector salud. También el grupo ha trabajado fuertemente en el diseño de sistemas digitales complejos basados en FPGAs para el procesamiento de señales como transformadas (Fourier, Wavelet, Hadamard, etc), procesadores de imágenes, sistemas de comunicaciones, sistemas criptográficos, etc; y en el diseño de sistemas electrónicos de señal mixta basados en FPAAs y PSoCs para instrumentación médica (ECG, monitor de signos vitales, y sistemas de fototerapia y magnetoterapia) e instrumentación industrial (medición de pérdidas en transformadores eléctricos, gases en alimentos y flujo en tuberías). Esta línea de desarrollo tecnológico nos permite tener la capacidad de diseñar los sistemas de instrumentación electrónica integrados en un solo chip para el procesamiento de las señales de los bionanosensores, microsensores y biochips (*embedded smart sensors* y *sensor arrays*).

Roles en la formulación del proyecto- Compromisos

-Formulación de la parte técnico-científica concerniente al Diseño computacional de bionanosistemas (bionanosensores, bionanodispositivos y bionanomáquinas) basados nanotubos de carbono, grafeno, polímeros o proteínas para diagnóstico y tratamiento médico usando biofármacos y vacunas moleculares.

-Formulación de los ejes articuladores del proyecto, considerando que es un proyecto multidisciplinario.

-Coordinador Logístico
-Participación en la redacción del marco lógico de la propuesta

Justificación de la asociación de los grupos

El objetivo es formar recursos humanos de alto nivel con el propósito de generar nuevo conocimiento que permita innovar en el diseño y fabricación de bionanosensores, microsensores y biochips para ser usados en aplicaciones que requieren el uso tecnología de punta, en especial, diagnóstico y tratamiento médico de enfermedades altamente prevalentes en nuestra región. Entonces, es necesario considerar una alianza estratégica con varios grupos de investigación para alcanzar los objetivos propuestos en el macroproyecto.

En este contexto, inicialmente el GBNE concentrará esfuerzos en el diseño y fabricación de bionanosensores, microsensores y biochips con base en nanotubos de carbono, grafeno, polímeros o proteinas para diagnóstico y tratamiento médico usando biofármacos y vacunas moleculares, es decir, el grupo se encargará de: a) desarrollar los modelos computacionales usando métodos como elementos finitos y dinámica molecular atomística; b) diseñar y simular los bionanosensores, microsensores y biochips usando los modelos computacionales; c) Fabricar los bionanosensores, microsensores y biochips usando la infraestructura de los laboratorios internacionales asociados al macroproyecto y considerando las técnicas del estado del arte en micro/nano fabricación.

Posteriormente, diseñar el hardware integrado para el procesamiento de la información generada por los bionanosensores, microsensores y biochips. En este caso, la primera versión del hardware será basada en los dispositivos reprogramables PSoCs/FPAAs, y luego el hardware será integrado en un solo chip para obtener los bionanosensores y microsensores inteligentes (*embedded smart-sensors*).

Finalmente, diseñar una arquitectura hardware usando los dispositivos reprogramables FPGAs para el secuenciamiento de ADN basado en el análisis multifractal. **- Asesores:** El centro contará con la participación de Asesores nacionales e internacionales en varias disciplinas: Los ingenieros John Sanabria del Grupo de redes y Víctor Manuel Vargas del Grupo AVISPA de la EISC-UniValle en computación distribuida y desarrollo de software, respectivamente, el Dr. José Manuel Gutiérrez en matemáticas y sistemas complejos de la Universidad de Cantabria, Santander, España y el Dr. Ashwin K. Naik MD y bioinformático de la India en salud pública, genómica y bioinformática.

-Grupo de Fisicoquímica de Bio y Nanomateriales

Coordinador: Rubén de Jesús Camargo.Universidad del Valle. Programa: Ingeniería Química
GrupLac: COL0015409201211139139

1. RESUMEN: Total numérico de cada aspecto.

Docentes Investigadores de tiempo completo (Por docentes de tiempo completo se entiende docentes que tienen una vinculación de tiempo completo con la institución proponente y que dedicarán parte de ese tiempo al programa propuesto).	**Producción Académica:** **(últimos 5 años)**	**Investigaciones** **(últimos tres años)**
Número de docentes: 3	Artículos internacionales: 1	Terminada: 6

Número de doctores: 3	Ponencias internacionales: 2	En ejecución: 3
Número de magísteres:	Artículos Nacionales: 5	
Otros (con otras titulaciones):	Ponencias Nacionales: 5	
	Patentes: 1	
Personal de apoyo:		
Número de doctores:	3	
Número de magísteres:		

LÍNEAS DE INVESTIGACIÓN DEL GRUPO (Se entiende por grupo de investigación una o más personas (investigador (es) con sus estudiantes y auxiliares) que se dedican a investigar conjuntamente una problemática particular en un área del saber. Según el tamaño del grupo y la complejidad de la problemática, éste puede desarrollar una o más líneas de investigación, correspondientes a aspectos específicos de dicha problemática. Dentro de cada línea de investigación, se pueden llevar a cabo uno o varios proyectos de investigación. En cada proyecto se puede desarrollar una o varias tesis de postgrado).

1. Electroquímica Básica y Aplicada
2. Fuentes de energía
3. Materiales para celdas de combustible
4. Materiales y Procesos para tratamientos en Salud

INFORMACIÓN DE ESTUDIANTES DE POSTGRADO: Número de estudiantes de Maestría que atiende el grupo: 5. Número de estudiantes de Maestría que podría atender el grupo. Número de estudiantes de Doctorado que atiende el grupo

Número de estudiantes de Doctorado que podría atender el grupo:

LISTADO DE PROFESORES INVESTIGADORES

NOMBRE	**TÍTULO MÁS ALTO (Institución que otorgó el título)**	**DEDICACIÓN (Dedicación del profesor a la institución y no al grupo)**
Rubén Jesús Camargo Amado	Doctorado Universidad Industrial de Santander – UIS, Doctorado en Ingeniería Química, 1999 - 2003.	Tiempo completo
William Hernando Lizcano Valbuena	Doutorado em Química - Instituto de Química de Sâo Carlos / Universidade de São Paulo - IQSC/USP - Brasil 2.003	Tiempo completo
Rubén Antonio Vargas Zapata	**Doctorado** University Of Illinois Doctor Of Philosophy In Physics Ph D de1973 - de 1977	Tiempo completo

-Grupo Estudios Doctorales en Informática – GEDI

El grupo se encuentra especializado en análisis masivo de información, computación distribuida y computación en la nube. Cuenta con varios estudiantes de pregrado y posgrado.

Las líneas de investigación en las cuales se especializa el grupo son Desarrollo de Software y tecnologías Web, Minería de datos, Multimedia y Visión, Tecnologías Computacionales Avanzadas, Tecnologías de la Información y las Comunicaciones. Lo cual tiene una amplia aplicación en el desarrollo de programas (software) y de actividades de base de datos.

Grupo de electroquímica

El grupo realiza actividades en el campo de las nanotecnologías y desarrollo de nanoproductos además del desarrollo de nuevos materiales. Tiempo: 2000-actual

Líneas de investigación
Análisis y remediación ambiental
Electroanalisis
Electroquimica
Espectroscopía
Sensores amperometricos de gases
Fotocatálisis
Polimeros conductores
Separaciones

El grupo cuenta con 29 investigadores, 37 artículos publicados y 17 proyectos desarrollados.

- Grupo de Biología Molecular, Ambiental y Cáncer, GBIMAC (Universidad del Cauca). Conformado por biólogos, químicos, médicos e ingenieros de sistemas, y Asesores Internacionales como: Dr. Ashwinikumar Naik Médico Bioinformático (Human Genome Project USA), Dr. José Manuel Gutiérrez - Universidad de Cantabria, España; los Drs. Felipe Rodríguez, Eduardo Fernandes EMBRAPA-Brasil y Jorge Martínez Queen University-Canada. Tiene 4 profesores PhDs: Maite Rada, Adriana Chaurra, Patricia Vélez y Pedro Moreno, 3 Magísteres, 2 candidatos a Doctor y varios estudiantes de doctorado, Maestría y Pregrado. Posee 15 publicaciones internacionales, 10 nacionales y varios artículos aun sometidos; 2 libros y un atlas en biología molecular. El grupo ha organizado (I y II Seminario Internacional de Genómica, Bioinformática, y Biología de Sistemas, 2000 y 2006) y participado en eventos internacionales; dirigido y codirigido más de 30 tesis de Pregrado, Maestría y varias tesis de doctorado en progreso. Ha contado con el apoyo de COLCIENCIAS, para financiación del proyecto "Análisis Multifractal del Genoma Humano" y el "Centro de Excelencia en Metagenómica y Bioinformática de Ambientes Extremos". Creó una vacuna sintética contra SIDA y un modelo teórico computacional y se diseñó un dispositivo para detectar SIDA por conducción eléctrica del DNA (Andrade *et al.*, 2012). Actualmente trabaja diseñando un microarray con aplicación multifractal para análisis de secuencias de DNA en patologías como cáncer, SIDA y diabetes y desarrollando un dispositivo con enfoque multifractal para diagnóstico fino de enfermedades cardiovasculares (Convocatoria Interna UniCauca, 2011). Estos desarrollos son aplicados para optimizar y personalizar los diagnósticos, tratamientos y/o control de enfermedades más comunes en Colombia.

-Grupo de Computación Científica -Centro de Bioinformática y Biología Computacional (Manizales)

Líneas de Investigación:
Algebra lineal numérica, Diseño geométrico asistido por computador,Estudio de superficies dadas por envolventes de esfera y plano, Aplicaciones geométricas a biología y química, Criptografía

El Grupo BIOS (o CBBC)
Director: Dago Hernando Bedoya Ortiz
Otros integrantes:
Cuenta con la asesoría de expertos internacionales, de países como Estados Unidos y España y muchos investigadores entre ellos Biólogos, Ingenieros de Sistemas y otros de la Universidad de Caldas, la Universidad Autónoma de Manizales, la Universidad de Manizales, la Universidad del Quindío, la Universidad Tecnológica de Pereira y la Universidad del Tolima, entre otros.

El Grupo BIOS se estableció mediante alianza público – privada liderada por el Ministerio de Tecnologías de la Información y las Comunicaciones (MinTIC) y el Departamento Administrativo de Ciencia, Tecnología e Innovación - Colciencias, en conjunto con Microsoft Colombia, Microsoft Research y un grupo de prestigiosas universidades de la zona cafetera, entre las cuales se encuentran: la Universidad de Caldas, la Universidad Autónoma de Manizales, la Universidad de Manizales, la Universidad del Quindío, la Universidad Tecnológica de Pereira y la Universidad del Tolima. Para la creación del centro el Ministerio TIC junto con Colciencias aporto recursos por $4.600 millones y las universidades de la región aportaron cerca de $400 millones que se reflejaron en la infraestructura física del centro. Por su parte, Microsoft hizo un aporte en cuanto a la arquitectura computacional, software de alto desempeño y las aplicaciones de los servicios que prestará el Centro con el objeto de generar desarrollo científico y de investigación para Colombia. El CBBC, por políticas establecidas por el Ministerio de las TIC permite a Colombia posicionarse como un país que potencia la utilización de TIC en áreas estratégicas como lo son la biotecnología y la biodiversidad, este centro de supercomputación central se establece para que una serie de centros de investigación especializados puedan funcionar en red a lo largo y ancho de la geografía colombiana. Con esta red, se pretende en la competitividad del país con base en su biodiversidad y en sus estudios de alta computación para otros campos como la medicina y la salud.
El CBBC cuenta con una capacidad de procesamiento de 145 Teraflops.

8. CRONOGRAMA DE ACTIVIDADES

1.	COMPONENTE	ACTIVIDADES	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
2.	1	Revisión bibliográfica	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		
3.		Identificación de pacientes y toma de muestras biológicas	x	X	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x												

4.		Almacenamiento de muestras biológicas	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x												
5.		Extracción de ADN	x	X	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x										
6.	2	Secuenciación exómico													x	x	x	x	x	x	x	x										
7.		Análisis de exomas																		x	x	x	x	x	x	x	x					
8.		Establecimiento de la infraestructura computacional		x	x	x	x	x	x	x	x	x	x	x																		
9.		Creación de un portal de información para la población experta (médica) y población en general			x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
10.		Software de análisis multifractal			x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x								
11.	3	Contratación de estudiantes de posgrado	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

12.	4	Preparación, escritura y sometimiento de artículos científico	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x				
13.	5	Diseño del diagrama de un chip potencial para diagnóstico								x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x						

9. PRESUPUESTO

Ver archivo Excel anexo.

REFERENCIAS

Adzhubei, Ivan A *et al.* (2010). A method and server for predicting damaging missense mutations. *Nature methods* 7(4): 248-249.

Banerji S, *et al.* (2012) Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature* 486(7403): 405–409.

Bluth MH, Hardin RE, Tenner S, Zenilman ME, Threatte GA. (2006).Laboratory diagnosis of gastrointestinal and pancreatic disorders. In: McPherson RA, Pincus MR, eds. *Henry's Clinical Diagnosis and Management by Laboratory Methods*. 21st ed. Philadelphia, Pa: Saunders Elsevier: chap 22.

Challis et al., 2012. An integrative variant analysis suite for whole exome next-generation sequencing data. *BMC Bioinformatics* 13:8.

Cifuentes Laura, Jara Lilian, Guillermo Barreto. (2013a). BRCA mutations and variants in low penetrance double-strands break repair genes and risk of breast cancer in South West Colombian population. *Breast Cancer Research.* Sometido.

Cifuentes, C. Cortés, A. Rivera, G. Barreto. (2013b). Identification of mutations associated with increased risk of hereditary breast and/or ovarian cancer in Colombia. *Eur. J Hum Genet* 21 (Suppl. 2): 528.

Cortés-Urrea, C., DM Tróchez-Jaramillo, M. Solarte-Cadavid, G. Barreto. (2012) Mutaciones en el exon 11 del gen BRCA1 y variantes en genes de baja penetrancia en pacientes con cáncer de mama familiar en Colombia. *Journal of Basic & Applied Genetics* 23:211.

Gracia-Aznarez F.J., Victoria Fernandez, Guillermo Pita, Paolo Peterlongo, Orlando Dominguez, Miguel de la Hoya, Mercedes Duran, Ana Osorio, Leticia Moreno, Anna Gonzalez-Neira, Juan Manuel Rosa-Rosa, Olga Sinilnikova, Sylvie Mazoyer, John Hopper, Conchi Lazaro, Melissa Southey, Fabrice Odefrey, Siranoush Manoukian, Irene Catucci, Trinidad Caldes, Henry T Lynch, Florentine S. M. Hilbers, Christi J. van Asperen, Hans F. A. Vasen1, David Goldgar, Paolo Radice, Peter Devilee, Javier Benitez. (2013). Whole Exome Sequencing SuggestsMuchofNon-BRCA1/ BRCA2 Familial Breast Cancer Is Due to Moderate and Low Penetrance Susceptibility Alleles. *PLOS ONE* 8(2) e55681.

Gilbert M.T., Haselkorn T., Bunce M., Sanchez J.J., Lucas S.B., Jewell L.D., Van Marck E., Worobey M (2007). The isolation of nucleic acids from fixed, paraffin-embedded tissues—which methods are useful when? *PLoS ONE* 2:e537.

Gillio-Tos A., De Marco L., Fiano V., Garcia-Bragado F., Dikshit R., Boffetta P., Merletti F. (2007). Efficient DNA extraction from 25-year-old paraffin-embedded tissues: study of 365 samples. *Pathology* 39: 345–348.

González, C. A., Navarro, C., Martínez, C., Quirós, J. R., Dorronsoro, M., Barricarte, A., Tormo, M. J., *et al.* (2004). Redalyc. El estudio prospectivo europeo sobre cáncer y nutrición (EPIC).

Häcker, G. (2000). The morphology of apoptosis. *Cell Tissue Res*. 301:5-14.

Hennekens CH., Buring JE. (1987). Epidemiology in Medicine Boston: Litte, Brown and Company.

Hidalgo-Miranda, A., Jiménez-Sánchez, G., Hidalgo-miranda, A., & Jiménez-Sánchez, G. (2009). Bases genómicas del cáncer de mama: avances hacia la medicina personalizada, 51.

Ivanov PCh, Nunes LA, Golberger AL, Havlin S, Rosenblum MG, Struzikk ZR,Stanley HE. (1999). Multifractality in human heartbeat dynamics. *Nature* 399:461-465.

Kelsey JL., Thompson WD., Evans AS. Methods in Observational Epidemiology. New York: Oxford University Press, 2da edición, Vol 26 (1996)

Kumar, Prateek, Steven Henikoff, and Pauline C Ng. (2009).Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature protocols* 4(7): 1073-1081.

Lanari, C., Molinolo, A. A., Viva, Q. (2003). Progestágenos y cáncer de mama: desarrollo de un modelo experimental. 2:111–121.

Ledford H. (2010). The cancer genome challenge. *Nature* 469: 972-974.

Li, H. *et al.* (2009). The sequence alignment/map format and SAM tools. *Bioinformatics* 25.16: 2078-2079.

Li, H. and Richard, D. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25.14: 1754-1760.

Li, Heng, Jue Ruan, and Richard Durbin. (2008). Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Research* 18.11: 1851-1858.

McCarthy, Mark I, Smedley, Damian and Hide, Winston. (2003). New methods for finding disease-susceptibility genes: impact and potential. *Genome Biology* 4:119

McKenna, Aaron *et al.* (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data *Genome Research* 20(9): 1297-1303.

McLeod HL (2013). Cancer pharmacogenomics: early promise, but concerted effort needed. *Science* 339:1563-1566.

Moreno PA. (2014). Bioinformática multifractal: una propuesta hacia la interpretación no lineal del genoma. *Ingeniería y Competitividad*. Revista científica y tecnológica. 16(1): xx - xx. *Aceptado.*

Moreno PA, Sandra Blanco, Patricia Eugenia Vélez Varela. (2006). Representación Secuencias de ADN y Proteínas mediante el juego del caos y su análisis multifractal. En: The 2nd International Seminar on genomics, Proteomics, Bioinformatics and Systems Biology, Universidad del Cauca, p.90 - 104 , v.1, fasc.1.

Moreno Pedro A, Vélez Patricia E, Martínez Ember, Garreta Luis E, Díaz Néstor, Amador Siler, Tischer Irene, Gutiérrez José M, Naik Ashwinikumar K, Tobar Fabián and García Felipe. (2011). The human genome: a multifractal analysis. *BMC Genomics* 12:506.

Pfeffer, U., *et al.* (2013). Breast Cancer Genomics: From Portraits to Landscapes. In Cancer Genomics (pp. 255-294). Springer Netherlands.

Puttonen, Katja A. *et al.* (2007). Different viabilities and toxicity types after 6-OHDA and Ara-C exposure evaluated by four assays in five cell lines. *Toxicology in Vitro* 30: 2-8.

Sanabria María Carolina, Gerardo Muñoz, Clara Inés Vargas. (2009). Análisis de las mutaciones más frecuentes del gen *BRCA1* (185delAG y 5382insC) en mujeres con cáncer de mama en Bucaramanga, Colombia. *Biomédica* 29: 61 – 72.

Sanabria, M. C., & Muñoz, G. (2009). Las mutaciones más frecuentes del gen BRCA1 (185delAG y 5382insC) en mujeres con cáncer de mama en Bucaramanga, Colombia; Mutations in the BRCA1.gene. Biomédica Bogotá, 29(1), 61–72. Retrieved fromhttp://redalyc.uaemex.mx/pdf/843/84311628009.pdf

Santisteban, A. S. (2006). Tema de Revisión Cáncer en el Siglo XXI Cancer in XXIst century, 23, 112–118.

Santos M.C., Saito C.P., Line S.R. (2008). Extraction of genomic DNA from paraffin-embedded tissue sections of human fetuses fixed and stored in formalin for long periods. *Pathol. Res. Pract.* 204: 633–636.

Santos S., Sa D., Bastos E., Guedes-Pinto H., Gut I., Gartner F., Chaves R (2009). An efficient protocol for genomic DNA extraction from formalin-fixed paraffin-embedded tissues. *Res. Vet. Sci.* 86: 21–426.

Sergio, A., Cuevas, R., & Rodríguez-cuevas, A. A. S. (2005). Cáncer de mama. 73: 423–424.

Shah SP, *et al.*(2012) The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* 486(7403):395–399.

Shi S.R., Datar R., Liu C., Wu L., Zhang Z., Cote R.J., Taylor C.R. (2004). DNA extraction from archival formalin-fixed, paraffin-embedded tissues: heat-induced retrieval in alkaline solution. *Histochem. Cell Biol* 122: 211–218.

Stephens PJ, *et al.* (2012). The landscape of cancer genes and mutational processes in breast cancer. *Nature* 486(7403): 400–404.

Torres A, Umaña A, Robledo J.F., Caicedo J.J., Quintero E., Orozco A., Torregrosa L, Tawil M., Hamman U & Briceño I. (2009). Estudio de factores genéticos para cáncer de mama en Colombia. Univ. Med. Bogotá, Colombia, 50 (3):297-301.

Torres D, Rashid MU, Gil F, Umaña A, Ramelli G, Robledo JF, *et al.*(2007). High proportionof BRCA1/2 founder mutations in Hispanic breast/ovarian cancer families from Colombia. *Breast Cancer ResTreat* 103: 225-32.

Vélez. PE. (1991). Estudio epidemiológico y etiológico del cáncer mamario humano, cultivos celulares in vitro de tumores primarios y efecto del 17b-estradiol y el tamoxifen sobre la proliferación celular. Tesis de Maestría. Universidad Nacional de Colombia. Bogotá, DC.

Vogelstein B, Papadopoulos N, Velculescu VE, Zhou Sh, Díaz LA Jr, Kinzler KW. (2013). Cancer Genome Landscape. *Science* 339: 1546-1558.

Wang, Kai, Mingyao Li, and Hakon Hakonarson.(2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research* 38.16: e164-e164.

Zografos G. Liakakos T, Roulos DH. (2013). Deep sequencing and integrative genome analysis: approaching a new class of biomarkers and therapeutic targets for breast cancer. *Pharmacogenomics* 14(1): 5-8.