



UNIVERSIDAD DE CHILE
Facultad de Ciencias
Escuela de Pregrado



UCH-FC
Biotecnología
G212
C.1

“ANÁLISIS DE VARIACIONES GENÉTICAS DEL LIGANDO MICA EN TUMORES DE PACIENTES CON ADENOCARCINOMA GÁSTRICO ”

Seminario de Título entregado a la Universidad de Chile en cumplimiento parcial de los requisitos para optar al título de Ingeniero en Biotecnología Molecular

VALENTINA CONSTANZA GÁRATE CALDERÓN

Director del Seminario de Título:
Dra. María Carmen Molina Sampayo

Co-Director del Seminario de Título:
Dr. Ricardo Amado Armisen Yañez

Patrocinante del Seminario de Título:
Dra. María Rosa Bono Merino

Abril, 2016
Santiago – Chile

ESCUELA DE PREGRADO – FACULTAD DE CIENCIAS – UNIVERSIDAD DE CHILE



INFORME DE APROBRACIÓN DE SEMINARIO DE TÍTULO

Se informa a la Escuela de Pregrado de la Facultad de Ciencias de la Universidad de Chile, que el Seminario de Título, presentado por la Srta. Valentina Constanza Gárate Calderón:

“ANÁLISIS DE VARIACIONES GENÉTICAS DEL LIGANDO MICA EN TUMORES DE PACIENTES CON ADENOCARCINOMA GÁSTRICO ”

Ha sido aprobado por la Comisión de Evaluación, en cumplimiento parcial de los requisitos para optar al Título de Ingeniero en Biotecnología Molecular.

Dra. María Carmen Molina Sampayo
Director Seminario de Título

Dr. Ricardo Armisen Yañez
Co-director de Seminario de Título

Dra. María Rosa Bono Merino
Patrocinante Seminario de Título

Comisión de Evaluación

Dr. Elías Utreras
Presidente Comisión

Dr. Miguel Allende
Evaluador

Santiago, Abril de 2016

Se muestran cuatro firmas manuscritas en azul sobre líneas horizontales. La segunda firma está sobre un sello circular que dice "FACULTAD DE CIENCIAS U. DE CHILE BIBLIOTECA CENTRAL".

*A mis padres, Mariela y Pablo,
Sin ustedes esto no habria sido posible.*

BIOGRAFÍA

Nací un 25 de Julio de 1990 en Santiago, hija de Pablo y Mariela, quienes cimentaron los valores de la persona que soy actualmente. Tuve una infancia feliz con mi hermana y muchos primos “hermanos de verano”, con los que compartí temporadas interminables en el campo. En mi etapa escolar fui muy autoexigente, siempre quise aprender más. Alrededor de los 13 años llegó a mis manos por casualidad una revista Bioplanet, leyendo esta revista descubrí lo que era la biotecnología y decidí que esta era la carrera que quería estudiar, por tal razón opté por cambiarme de colegio a Santiago para tener más posibilidades de entrar a la Universidad. Fue así como en 1ro Medio ingresé al Liceo Carmela Carvajarl de Prat y me vida cambió en 180 grados. Finalmente encontré gente con la que podía debatir y mi mente se abrió a otro mundo. Ya no era la primera de la clase; y además logré forjar amistades que han probado ser duraderas.

Me licencié de 4to Medio dentro del 5% superior de mi generación, rendí la PSU y logré cumplir mi sueño de ingresar a la carrera de Ingeniería en Biotecnología Molecular, de la Universidad de Chile, con el fin de aplicar mis conocimientos a favor del desarrollo de mi país a través de la Ciencia.

Actualmente estoy realizando el Magíster en Cs. Biológicas en la Facultad de Medicina de la Universidad de Chile, el cual me entrega herramientas para emigrar y seguir cumpliendo mis metas profesionales.

AGRADECIMIENTOS

En primer lugar, quiero agradecer a la Dra. María Carmen Molina y en especial al Dr. Ricardo Armisen por la oportunidad que me dieron de realizar este trabajo de investigación bajo su tutela. Esta instancia me dio la oportunidad de aprender como nunca sobre el quehacer científico y plantearme nuevos desafíos.

Le agradezco a Marce y Jessy (el Team de Secuenciación), con quienes tuve la oportunidad de compartir esta experiencia de principio a fin, por su amistad y buena onda. A los chicos del IBT, Norbert, Marti y Roberto, por su constante apoyo en los primeros pasos como una científica de verdad. A Lorenz, Kari, Maca, Caro, Felipe, Jaime, Gabriel, Leo, Bastián, Gerardo, Eduardo, por su buena disposición y alegría que le dieron al laboratorio.

A muchos otros integrantes del PDI y del RA lab, como Profe Caro, Sra. Nancy, Inelia, Perry, Dany, Fer, Tutu, Eduardo y Alfredo, por hacer que mi estadía en el laboratorio fuera muy grata.

Agradezco a mis compañeros Biotecs, especialmente a Nia, Luisfe, Cami, Pito, Sama, Ceci, Sergio, Arantza, Kiki, Mami, Silvi, Camilo, Herminito (Pablo Flores), Carlitos, Pablo Mancilla, Erick, Óne-o y al Papi, porque han estado en muchos momentos dándome fuerzas para terminar este desafío y entregarme lindos recuerdos de esta aventura.

A mis Fresccas de la vida, Maca, Pazi, Gina, Anto, Cher y Milay por todos esos buenos momentos, conversaciones sobre la vida, por su compañía y amistad infinita.

A mi primo Omar, quien ha sido mi hermano mayor, mi partner y mi mejor amigo; por su apoyo incondicional desde siempre.

Finalmente, agradecer a Dios y a mi familia. Mi hermana Paulina, mis papás Pablo y Mariela, por su paciencia y el amor infinito que he recibido. Gracias a ustedes me he convertido en la mujer que soy hoy. Este trabajo es para ustedes.

FINANCIAMIENTO

Este trabajo se realizó en el laboratorio de Inmunovigilancia y Evasión Inmune del Programa de Inmunología, en colaboración con el laboratorio de Patología Molecular del Cáncer, ambos ubicados en el instituto de Ciencias Biomédicas (ICBM) de la facultad de Medicina de la Universidad de Chile. Su realización fue posible con el financiamiento de los proyectos FONDECYT 1130330 (M.C.M), y 1151446 (R.A), FONDEF D1111029 (R.A) y ANILLO ACT1115 (R.A).

ÍNDICE

BIOGRAFÍA	iv
AGRADECIMIENTOS	v
FINANCIAMIENTO	vii
ÍNDICE	viii
ÍNDICE DE TABLAS	x
ÍNDICE DE FIGURAS	xi
LISTA DE ABREVIATURAS	xiii
RESUMEN.....	xv
ABSTRACT	xvii
INTRODUCCIÓN	1
1.1 Cáncer Gástrico	1
1.2 Sistema Inmune y células Natural Killer.....	2
1.3 Activación de las células NK	5
1.4 Receptor NKG2D y sus ligandos	6
1.5 Moléculas NKG2D y sus ligandos en la inmunidad antitumoral	7
1.6 Variaciones genéticas de <i>MICA</i> como mecanismos de evasión de inmunidad antitumoral.....	10
1.7 Hipótesis.....	15
1.8 Objetivo General	15
1.9 Objetivos Específicos.....	15
2. MATERIALES Y MÉTODOS	16
2.1. Pacientes	16
2.2. Análisis de calidad del material genético extraído desde muestras.....	18
2.2.1. Preparación del tejido tumoral	18
2.2.2. Extracción del ADN genómico	18
2.2.3. Cuantificación de ADN mediante densidad óptica	18
2.2.4. Cuantificación de ADN de doble hebra	19
2.2.5. Electroforesis de ADN en geles de agarosa	20
2.2.6. Electroforesis Capilar de ADN	21

2.3. Secuenciación masiva de exones del gen <i>MICA</i>	21
2.3.1. Diseño de amplicones	21
2.3.2. Preparación de bibliotecas.....	22
2.3.3. Secuenciación en equipo MiSeq	23
2.4. Análisis Bioinformático.....	26
2.4.1. Comprobación de la calidad de los datos.....	26
2.4.2. Alineación de las lecturas con el genoma de referencia.	30
2.4.3 Localización de las variantes presentes en el gen <i>MICA</i>	32
2.4.4. Filtrado de las variantes detectadas.....	32
2.4.5 Identificación de mutaciones somáticas.....	37
2.4.6. Predicción del efecto de mutaciones somáticas en la proteína <i>MICA</i>	37
2.4.7. Visualización de los resultados.	38
3. RESULTADOS.....	40
3.1 Selección de muestras en estudio y secuenciación masiva de exones del gen <i>MICA</i>	40
3.1.1 Integridad del material genético.....	40
3.1.2 Diseño de amplicones.	49
3.1.3 Métricas de la secuenciación.....	53
3.2 Detección de variantes en el gen de <i>MICA</i>	59
3.2.1 Identificación de variantes genéticas en las muestras en estudio.....	59
3.2.2 Caracterización de nuevas variantes del gen <i>MICA</i>	61
3.2.3 Localización y distribución de las variantes genéticas en el gen <i>MICA</i>	68
3.3 Estudio de mutaciones somáticas en el gen <i>MICA</i> y determinación de su efecto en la proteína.	70
3.3.1 Identificación de las mutaciones somáticas en el gen <i>MICA</i>	70
3.3.2 Localización de las mutaciones somáticas.....	74
3.3.3 Predicción del efecto de las mutaciones somáticas en la proteína <i>MICA</i>	76
4. DISCUSIÓN	80
4.1 Detección de variaciones en el gen <i>MICA</i>	81
4.2 Mutaciones somáticas en el gen <i>MICA</i> y su efecto en la proteína	86
5. CONCLUSIONES	93
6. ANEXOS.....	94
7. BIBLIOGRAFÍA.....	98

ÍNDICE DE TABLAS

Tabla 1: Características clínico-patológicas de los pacientes con cáncer gástrico.....	17
Tabla 2: Protocolo preparación curva estándar para ensayo PicoGreen.....	19
Tabla 3: Lista de filtros utilizados en la detección de variantes.....	34
Tabla 4: Detalle de la cuantificación de cada muestra analizada.....	46
Tabla 5: Efectos de las nuevas variantes en la proteína MICA	67
Tabla 6: Frecuencia alélica promedio de las mutaciones somáticas detectadas.....	71
Tabla 7: Mutaciones somáticas para el gen <i>MICA</i> detectadas en las muestras en estudio	73

ÍNDICE DE FIGURAS

Figura 1 Organización exón-intrón del gen <i>MICA</i>	11
Figura 2 Flujo de trabajo durante la secuenciación masiva en el equipo MiSeq	25
Figura 3 Representación de una lectura del secuenciador en formato FASTQ.....	27
Figura 4 Gráficos de parámetros de calidad de los datos generados con el programa FastQC.....	29
Figura 5 Herramientas informáticas utilizadas en el análisis de los datos	31
Figura 6 Visualización de archivo VCF en Excel	36
Figura 7 Ejemplo de visualización de los datos con el programa IGV.....	39
Figura 8 Agrupación de muestras en estudio de acuerdo a la integridad del material genético	41
Figura 9 Distribución de fragmentos de ADN genómico en muestras del Grupo I y II	43
Figura 10 Distribución de fragmentos de ADN genómico en muestras del Grupo III y IV	44
Figura 11 Selección de las muestras a ser secuenciadas	48
Figura 12 Visualización de regiones adicionales en el estudio del gen <i>MICA</i>	50
Figura 13 Visualización de los amplicones diseñados versus los entregados por Illumina Inc. para los exones en estudio del gen <i>MICA</i>	52
Figura 14 Porcentaje de lecturas identificadas que pasaron el filtro de calidad de Illumina para cada una de las muestras de cáncer gástrico secuenciadas.....	54
Figura 15 Puntuación de calidad (Q Score) del llamado de bases para todos los ciclos de secuenciación.....	56

Figura 16 Cobertura promedio por amplicón para el gen <i>MICA</i>	58
Figura 17 Distribución de las variaciones genéticas para el gen <i>MICA</i> detectadas por secuenciación masiva en pacientes con adenocarcinoma gástrico.....	60
Figura 18 Localización de nuevas variantes detectadas en el gen <i>MICA</i>	62
Figura 19 Visualización de nueva variante de <i>MICA</i> (NM_000247:c.944T>G; p.Val302Gly) identificada por secuenciación masiva.....	64
Figura 20 Visualización de nueva variante de <i>MICA</i> (NM_000247:c.946C>A;p. Leu303Ile) identificada por secuenciación masiva	65
Figura 21 Distribución de los SNV detectados para el gen <i>MICA</i> e identificación de variantes presentes en una base de datos chilena	69
Figura 22 Localización de mutaciones somáticas detectadas para el gen <i>MICA</i>	75
Figura 23 Efecto de la mutación P294 en la estructura de la proteína <i>MICA</i>	77
Figura 24 Visualización de la mutación P294A en el dominio alfa 3 de la proteína <i>MICA</i>	79

LISTA DE ABREVIATURAS

ADN	Ácido Desoxirribonucleico
ADNg	ADN genómico
ARN	Ácido Ribonucleico
ARNm	ARN mensajero
CG	Cáncer gástrico
Chr	Cromosoma
dbSNP	Base de datos de polimorfismos de sólo nucleótido
dNTP	Desoxinucleótido trifosfato
GWAS	Estudio de asociación del genoma completo
Hg19	Genoma de referencia humano versión 19
IAS	Servicio de Anotación de Illumina
IFN-γ	Interferón gamma
Ig	Inmunoglobulina
IL	Interleuquina
IMGT	Base de datos ImmunoGenetics
kDa	Kilodalton
MHC-I	Complejo principal de histocompatibilidad de clase I

MICA/B	Proteína relacionada a cadena A/B de la molécula de clase I del MHC
NK	Natural Killer
NKG2D	Receptor de activación de células NK, grupo 2, miembro D
NKG2DL	Ligando de NKG2D
NKT	Natural Killer T
Pb	Pares de bases
PCR	Reacción en cadena de la Polimerasa
RAET	Transcritos inducidos por ácido retinoico
SIFT	Sorting intolerant from tolerant
sMICA	MICA soluble
SNP	Polimorfismo de un sólo nucleótido
SNV	Variante de un solo nucleótido
SVC	Somatic Variant Caller
TGF-β	Factor de crecimiento transformante beta
TNF-α	Factor de necrosis tumoral alfa
ULBP	Proteínas de unión a UL16
VCF	Variant call format

RESUMEN

El Cáncer Gástrico tiene una alta incidencia en Chile constituyendo la primera causa de muerte por tumores malignos. La respuesta inmune innata es considerada como la primera línea de defensa tumoral, siendo uno de sus componentes más importantes las células Natural Killer (NK), las cuales están especializadas en la eliminación de células tumorales o infectadas por virus. La actividad de las células NK está regulada por un complejo balance de receptores inhibidores y activadores. Entre los receptores de activación se destaca NKG2D, el que se expresa en células NK y linfocitos T. Se han descrito diferentes ligandos de NKG2D (NKG2DL) como MICA, MICB y ULBP1-6. Estos marcadores presentan expresión restringida en células normales y una inducción en respuesta al estrés celular y transformación tumoral, permitiendo la eliminación de las células estresadas por el sistema inmune. Este proceso se desarrolla mediante la interacción NKG2D-ligando, lo que gatilla una señal de activación en la célula que expresa NKG2D, promoviéndose lisis citotóxica de la célula que expresa el ligando. Se han reportado diversos polimorfismos en el gen *MICA* que estarían implicados en su interacción con el receptor y que darían cuenta de mecanismos de evasión de la respuesta inmunitaria durante la progresión tumoral. En consecuencia, esta memoria se propone que durante el desarrollo del cáncer gástrico se originan mutaciones en el gen que codifica *MICA*, las que estarían afectando la interacción ligando-receptor y promoviendo la evasión inmune por parte de las células tumorales. El objetivo de este trabajo fue identificar variantes somáticas de nucleótido simple (SNV) en el gen *MICA* en pacientes chilenos con cáncer gástrico y analizar los efectos de estos cambios en la proteína. En este estudio se incluyeron cincuenta pacientes con adenocarcinoma gástrico

del Departamento de Cirugía Digestiva del Hospital del Salvador. La detección de mutaciones en la secuencia de *MICA* se realizó por secuenciación masiva en la plataforma MiSeq de Illumina. Para el análisis bioinformático de las secuencias se empleó el algoritmo Somatic Variant Caller (Illumina). Detectándose 3 variantes nuevas, no descritas previamente en bases de datos, de las cuales debido al puntaje entregado por el algoritmo SIFT sólo la variante chr6: 31380116 tendría un efecto en la proteína MICA.

Por otra parte, en base a las frecuencias alélicas se identificaron 6 mutaciones somáticas de tipo SNV. Sin embargo, en relación a los datos entregados por la base de datos SNPeffct sólo la variante P294A presentó una reducción en la estabilidad de la proteína MICA. A su vez, la localización de este cambio se presentó en una región de posibles cortes proteolíticos mediados por metaloproteasas, sugiriendo que esta mutación podría estar dando cuenta de la liberación de ligandos solubles por parte de la célula tumoral, promoviendo un mecanismo de evasión al sistema inmune.

Nuestros resultados sugieren que la presencia de SNVs en el gen *MICA*, en pacientes con adenocarcinoma gástrico, podría corresponder a una estrategia de evasión inmune que esté favoreciendo el establecimiento y desarrollo del tumor.

ABSTRACT

Gastric cancer (GC) is the principal causes of cancer mortality in Chile. Innate immune response is considered the first tumoral defense line, indeed, one of its most important components are Natural Killer (NK) cells, which are specialized in the tumoral cells elimination or viruses infected cells. The activity of NK cells is regulated by a complex balance of inhibitors and activators receptors. The activating receptor NKG2D, whose ligands (NKG2DL), expressed on the T lymphocytes and NK cell surface, belong to the MIC (MICA and MICB) and ULBP/RAET (ULBP1-6) families. NKG2D have restricted expression in normal cells and an induction in response to cellular stress and tumor transformation, allowing the removal of stressed-out cells by the immune system. This process develops through the interaction NKG2D-ligand, which trigger an activation signal in the cell that expresses NKG2D, promoted the cytotoxic lysis in the cell that expresses the ligand. Different *MICA* polymorphisms indicate this gene would be involved in the interaction with the receptor and would explain mechanisms of tumor evasion, during tumoral progressing. In summary, in this memory propose that during the development of gastric cancer, take place mutations in the gene that encodes MICA protein, which would affect the ligand-receptor interaction and promoting the immune evasion by tumoral cells. The aim of this study was to identify somatic single nucleotide variants (SNV) in the *MICA* gene, in Chilean patients with gastric cancer and analyze the effects of these changes in the protein. Were included in this study fifty patients with gastric adenocarcinoma of the Department of Digestive Surgery of the Hospital El Salvador and the detection of mutations in the sequence of *MICA* was performed by massive sequencing in the MiSeq platform of Illumina. The bioinformatic analysis of

sequences was performed with the algorithm Somatic Variant Caller (Illumina), three new variations were detected, this had not been previously described in databases, of which due to the score given by the SIFT algorithm only the variant chr6: 31380116 would have an effect on the MICA protein. Also, were identified 6 SNV somatic mutations according to allele frequencies. However, in relation to the data released by the database SNPeffect only variant P294A presented a reduction in the MICA protein stability. The location of this change is introduced in a region of possible proteolytic cuts by metalloproteases, suggesting that this mutation could be giving the release of soluble ligands by the tumor cell, promoting an evasion mechanism of the immune system. Our results suggest that the presence of SNVs in the *MICA* gene, in patients with gastric adenocarcinoma, it might be related with immune evasion strategy which provides establishment and development of tumor.

1. INTRODUCCIÓN

1.1 Cáncer Gástrico

El cáncer es un proceso de crecimiento y propagación incontrolada de células, generado por un desorden hiperproliferativo donde las células adquieren la capacidad de ser autosuficientes en las señales de crecimiento y ser insensibles a las señales que lo inhiben. Estas células cancerosas tienen potencial proliferativo ilimitado, pueden evadir la apoptosis, estimular la angiogénesis sostenida, evadir la destrucción inmune adquiriendo la capacidad de invadir tejidos y generar metástasis en puntos distantes del organismo (Hanahan & Weinberg, 2011).

El Cáncer Gástrico (CG) es el cuarto más común en el mundo, con una incidencia del 8% (con 951.600 casos nuevos) y representa la segunda causa de muerte relacionada con cáncer, abarcando el 10% de éstas (723.100 muertes por año) (Torre y col., 2015). Chile se encuentra entre los países con las tasas de mortalidad más altas junto a Singapur, Costa Rica y Japón (Lee y col., 2006), representando la primera causa de muerte por tumores malignos para ambos sexos (MINSAL, 2010), con una tasa de mortalidad en torno a 20 por 100.000 habitantes, falleciendo alrededor de 3.000 personas al año.

El CG es una enfermedad multifactorial gatillada de la interacción entre la susceptibilidad genética propia del individuo y factores ambientales (McLean, 2014). De estos últimos, uno de los factores principales es la dieta; el consumo de alimentos que contienen nitritos, alimentos ahumados y/o salados aumenta el riesgo de desarrollar esta enfermedad. Además la infección por *Helicobacter pylori* (*H. pylori*) se ha

identificado como una de las causas del desarrollo de esta patología (Kandulski y col., 2010).

El CG agrupa tumores que se encuentran bajo la unión gastroesofágica, correspondiendo generalmente a un adenocarcinoma. En la mayoría de los casos progresa de manera lenta, asintomática y en nuestro país, más de la mitad de los pacientes se encuentran en un estado avanzado de la enfermedad al momento del diagnóstico (MINSAL, 2010). Los tratamientos disponibles para pacientes con CG incluyen: cirugía, quimioterapia, radioterapia o una combinación de ellas. Sin embargo, el único tratamiento potencialmente curativo corresponde a la gastrectomía total o parcial acompañada con linfadenectomía (Okines y col., 2010); no obstante, de los pacientes intervenidos un 40% a 65% de ellos podrían presentar recurrencia de la enfermedad (Dicken y col., 2005). Los pacientes tratados tienen una esperanza de vida promedio de 24 meses y sólo un 10% a 30% una sobrevida de 5 años. Teniendo en cuenta la alta incidencia del CG en Chile, las poco esperanzadoras estadísticas en cuanto a resultados de tratamientos y las exitosas estrategias preventivas de países como Japón (Asaka y col., 2014), es que se considera impostergable incluir dentro de las políticas públicas la prevención y detección precoz del CG en nuestro país.

1.2 Sistema Inmune y células Natural Killer

El sistema inmune en mamíferos es una compleja red de moléculas, células y órganos que interactúan para proteger al organismo tanto del daño por elementos externos (provocado por agentes patógenos), así como por agentes internos (tales como células neoplásicas). Cabe señalar, que esta función es realizada sin ocasionar daño a las células

propias del organismo, debido a que este sistema posee la capacidad de discriminar lo que es propio de lo que no lo es a través del reconocimiento de moléculas específicas (Visser & Coussens, 2006).

El desarrollo de la fase efectora del sistema inmune implica procesos de tolerancia, los que están balanceados por complejos mecanismos de control que favorecen su acción efectora e inhibitoria. En este contexto, se ha demostrado que agentes patógenos y neoplasias han desarrollado estrategias que les han permitido evadir los sistemas de control de la respuesta inmune, evitando su reconocimiento y eliminación, fenómeno conocido como Evasión Inmune (Eagle & Trowsdale, 2007).

Por su parte, el sistema inmune Innato está compuesto por células circulantes y proteínas plasmáticas, cuyas funciones son: la protección del organismo frente a agentes agresores externos, el control de procesos infecciosos y la defensa frente a agresores endógenos (como neoplasias), participando en el control de los procesos antitumorales y antimetastásicos (Karre y col., 2005). La activación del sistema inmune Innato trae consigo la respuesta inmune adaptativa mediada por los linfocitos T y B. Estas células portan receptores antigénicos que permiten distinguir antígenos propios de los exógenos y dirigir eventos posteriores. Las células tumorales son propias del organismo y difieren sólo sutilmente en su comportamiento y patrón bioquímico. Por tal razón, la Inmunidad Innata al ser la primera barrera de defensa contra estas células, tiene un rol efector fundamental en las primeras etapas del desarrollo del cáncer, proceso denominado Inmunovigilancia (Dunn y col., 2004).

Uno de los componentes más importantes del sistema inmune innato son las células Natural Killer (NK) (células citolíticas naturales), las cuales corresponden a la primera línea de defensa contra ciertas infecciones virales y tumores (Trinchieri, 1989). Cabe destacar que estas células son capaces de reconocer y lisar células tumorales e infectadas con virus sin sensibilización previa (Cerwenka & Lanier, 2001).

Las células NK son linfocitos derivados de la médula ósea, y la mayoría se encuentra localizada en la sangre periférica, linfonodos, bazo y médula ósea (Ferlazzo y col., 2004), y en humanos, comprenden entre un 5-20% de los linfocitos de sangre periférica.

Las células NK eliminan las células blanco (células tumorales o células infectadas por virus) mediante diversos mecanismos, uno de ellos corresponde a la liberación de gránulos citoplasmáticos que contienen proteínas como perforina y granzimas (Trapani & Smyth, 2002), las cuales promueven lisis celular por osmosis e inducen apoptosis de las células blanco, respectivamente (Podack & Dennert, 1983; Dennert y col., 1987).

Una alternativa a la liberación de gránulos es la muerte de la célula blanco inducida por apoptosis mediante receptores de muerte presentes en estas. Las células NK pueden expresar ligandos de muerte de la familia del factor de necrosis tumoral (TNF), como ligando Fas, TNF- α y TRAIL (ligando inductor de apoptosis relacionado al TNF), los cuales se unen a receptores específicos en la superficie de la célula blanco, gatillando el proceso de apoptosis (Zamai y col., 1998).

Otro mecanismo de acción consiste en la secreción de citoquinas inmunoregulatoras, siendo capaces de expresar interferón (IFN)- γ , TNF- α , TNF- β , factor estimulante de

crecimiento de granulocitos y macrófagos (GM-CSF), interleuquina (IL)- 10 (Deniz y col., 2008) e IL-13 (Loza y col., 2002). Dentro de estas citoquinas el IFN- γ es de gran importancia en la inmunidad antitumoral, ya que inhibe la proliferación celular, restringe la angiogénesis tumoral, estimula la presentación de antígenos, promueve la apoptosis y estimula la respuesta inmune adaptativa, entre otros roles (Dunn y col., 2006).

1.3 Activación de las células NK

Las células NK expresan en su superficie un amplio repertorio de receptores activadores e inhibitorios, los que se unen a moléculas presentes en la superficie de células blancas tales como MHC-I (Complejo principal de histocompatibilidad de clase I) y moléculas relacionadas con MHC-I. El reconocimiento de las células blancas por las células NK involucra una alteración en el balance entre señales activadoras e inhibitorias. Estas señales son entregadas simultáneamente a las células NK por distintas familias de receptores presentes en su superficie y que se unen a ligandos expresados por las células blancas (Smyth y col., 2005). Sin embargo, la pérdida de la molécula MHC-I en la superficie de la célula blanca, ya sea por mutaciones genéticas que ocurren durante el desarrollo tumoral o por mecanismos de evasión viral; eliminaría los controles inhibitorios, predominando las moléculas activadoras y resultando en la activación de la célula NK. Este fenómeno es conocido como Reconocimiento por Ausencia de lo Propio (Shifrin y col., 2014) y el resultado de esta interacción va a determinar si la célula NK es finalmente activada, si se producen citoquinas y/o si la célula blanca es eliminada (Caligiuri, 2008).

1.4 Receptor NKG2D y sus ligandos

El receptor NKG2D (Natural Killer Group 2, type D) se expresa en todas las células NK y ha sido identificado en linfocitos T $\alpha\beta$ CD8+, T $\gamma\delta$ y NKT (Chan y col., 2006). Este receptor es un homodímero que pertenece a la familia de los receptores tipo lectinas del tipo C (Houchins y col., 1991) y consiste en dos proteínas de membrana de tipo II unidas por un puente de disulfuro. El dominio de membrana posee aminoácidos cargados positivamente y el dominio intracelular es muy corto y no tiene la propiedad de señalización intracelular. En humanos, el receptor NKG2D forma un complejo con la molécula adaptadora DAP10 (proteína activadora DNAX de 10 kDa), la cual se asocia con el receptor NKG2D como homodímeros (Wu y col., 1999). DAP10 tiene en su cola citoplasmática un motivo Tyr-Ile-Asn-Met (YINM), que al ser fosforilado genera una señal que estimula la proliferación, citotoxicidad de las células NK y provee de co-estimulación a los linfocitos T activados (Chang y col., 1999; Wu y col., 2000).

En humanos se han descrito dos familias de ligandos para NKG2D (NKG2DL):

A) La familia de proteínas relacionadas con la cadena de MHC de clase I (MIC). En este complejo se encuentran codificadas: MICA y MICB. Los que presentan similitud a la cadena α del MHC-I con los dominios $\alpha 1$, $\alpha 2$ y $\alpha 3$, pero se diferencian de éstos al no unir $\beta 2$ - microglobulina ni presentar antígenos (Zwimer y col., 1998).

B) La familia de proteínas de unión a la molécula UL16 del citomegalovirus, la cual consta de 6 miembros (ULBP1-6) (Cosman y col., 2001). Sin embargo, debido a que sólo ULBP 1 y 2 unen realmente a la proteína UL16, la denominación actual es

Transcritos Inducidos por Ácido Retinoico 1 (RAET1: RAET1I (ULBP1), RAET1H (ULBP2), RAET1N (ULBP3), RAET1E 1 y 2 (ULBP4), RAET1G (ULBP5), RAET1L (ULBP6) (Chalupny y col., 2003; Eagle y col., 2009). Estas moléculas son similares a los ligandos MIC, pero presentan sólo los dominios $\alpha 1$ y $\alpha 2$ de la cadena α y no se codifican dentro del complejo MHC.

Los NKG2DL no se expresan comúnmente en la superficie en tejidos normales. Sin embargo, su expresión en superficie aumenta en diversos tipos celulares bajo condiciones de estrés (shock térmico), infección y frecuentemente durante la progresión tumoral. En este punto, es importante recalcar que debido a que los tumores se desarrollan a partir de células propias, son generalmente poco inmunogénicos y muchas veces no son reconocidos eficientemente por el sistema inmune, por tal razón la presencia de estos ligandos que se expresan en células y líneas tumorales (Nausch, 2008) servirían para poder detectar las células transformadas. Estudios previos han establecido que la expresión de NKG2DL en tumores los hace susceptibles a su eliminación por las células NK *in vitro* y además la formación de tumores puede ser prevenida a través de la señalización por NKG2D (Gonzalez y col., 2006).

Por lo tanto, el receptor NKG2D juega un rol importante en la Inmunidad antitumoral a través de la interacción entre el sistema inmune del huésped y el tumor.

1.5 Moléculas NKG2D y sus ligandos en la inmunidad antitumoral

La inmunidad antitumoral se puede describir en tres fases: Eliminación, Equilibrio y Evasión, las que se describen en la Teoría de las 3 E de la Inmunoección del cáncer

(Dunn y col., 2004). Durante la Eliminación, el sistema inmune del huésped puede prevenir un crecimiento tumoral, trabajando en conjunto con la respuesta inmune innata y adaptativa sin que haya progresión a las siguientes fases mediante la liberación de citoquinas proinflamatorias, las que reclutan células del sistema inmune innato como: macrófagos, células dendríticas y células NK. En tanto, durante la fase de Equilibrio debido a la rápida tasa de reproducción y mutación que tienen las células tumorales, se produce una selección de clones resistentes al sistema inmune del huésped. Este estado se puede mantener durante muchos años, en donde las células tumorales son eliminadas mientras los clones resistentes permanecen en los tejidos (Burnet, 1971). Durante la fase final de la Inmunoedición que corresponde a la Evasión, los tumores desactivan el reconocimiento inmune permitiendo la progresión de la enfermedad, promoviendo la angiogénesis y la invasión de tejidos, siendo capaz de sobrepasar la respuesta inmune y desencadenar metástasis.

El escape inmune y la progresión tumoral constituyen distintos mecanismos inmunosupresores, los que pueden incluir la pérdida de la maquinaria de presentación de antígenos, expresión de moléculas inhibitorias de apoptosis e inducir células T reguladoras (Dunn y col., 2004).

En cuanto a los mecanismos de Evasión de la Inmunovigilancia, existe evidencia de que a pesar de la expresión de NKG2DL en tumores y que tales moléculas se presentan como señales de muerte, informándole a las células NK que están presentes y gatillando el ataque citolítico, en la mayoría de los casos el tumor logra sobrevivir y escapar del ataque del sistema inmune. Teniendo en cuenta esta contradicción, se ha planteado que

podría existir una selección negativa sobre las células tumorales que expresan NKG2DL en su superficie y en particular investigaciones en melanoma, indican que la expresión de estos ligandos depende del estado de avance tumoral, existiendo una alta expresión de estos ligandos en las lesiones primarias, pero disminuye en las lesiones metastásicas. Por lo que con el tiempo quedarían sólo las células tumorales menos inmunogénicas, es decir aquellas que no expresan o poseen una baja expresión de los NKG2DLs (Vetter y col., 2004). Otro mecanismo que se ha documentado es una expresión continua de MICA, unido a la membrana, regulando negativamente la activación de las células NK por el receptor NKG2D, lo que compromete la capacidad de las células efectoras en reconocer y consecuentemente, lisar las células tumorales. Por otra parte, se han detectado formas solubles tanto de ligandos MIC como RAET1, lo cual tendría relación con la disminución de la cantidad y funcionalidad del receptor NKG2D en la superficie de las células por la internalización y la degradación lisosomal de NKG2D, disminuyendo así sus niveles de expresión en superficie en linfocitos T CD8+ y células NK, afectando la citotoxicidad (Groh y col., 2002). Adicionalmente, se ha reportado que formas solubles de MIC-A/B (sMIC) también desactivarían la inmunidad de las células NK, debido a que pueden actuar como inhibidor competitivo bloqueando el reconocimiento de moléculas MIC unidas a la membrana. Por otra parte, la expresión de NKG2DL solubles se debe a que pueden ser clivados por metaloproteinasas secretadas por células tumorales, lo que conlleva en la liberación de formas solubles de los ectodominios de los ligandos de NKG2D (Coudert y col., 2006). Es así como se han detectado ligandos solubles en muestras serológicas de pacientes de distintos tipos de cáncer; en particular, sMICA en el suero de pacientes con cáncer de próstata, cáncer de colon, cáncer

páncreas, neuroblastoma, osteosarcoma y tumores hematopoyéticos severo (Tamaki y col., 2009). Por tal razón, aunque los tumores expresan NKG2DL, estos crecen progresivamente en individuos sanos debido a los múltiples mecanismos de evasión de la respuesta inmune antes presentados.

El hecho de que las células clonales tumorales evadan los mecanismos de control ejercidos por las células NK, puede deberse a la ocurrencia de mutaciones en el ADN de las células neoplásicas que conlleven a la alteración en la interacción de NKG2D con sus ligandos, desde el punto de vista de su afinidad o expresión del ectodominio que permita la liberación de los ligandos al plasma. En este contexto, se han reportado numerosas variaciones en las secuencias de los NKG2DL de tipo polimórficas que podrían estar implicadas en la evasión (Antoun y col., 2010).

1.6 Variaciones genéticas de *MICA* como mecanismos de evasión de inmunidad antitumoral

El gen *MICA* es transcrito en un ARNm de 1.382 pb, originando un polipéptido de 383 aminoácidos con un peso molecular de 43 kDa (Bahram y col, 1994). La proteína MICA es altamente glicosilada, ya que tiene 8 sitios potenciales de N-glicosilación ubicados a lo largo de sus 3 dominios extracelulares, por lo que la proteína madura tiene un peso molecular 65 kDa (Groh y col, 1996) (Figura 1).

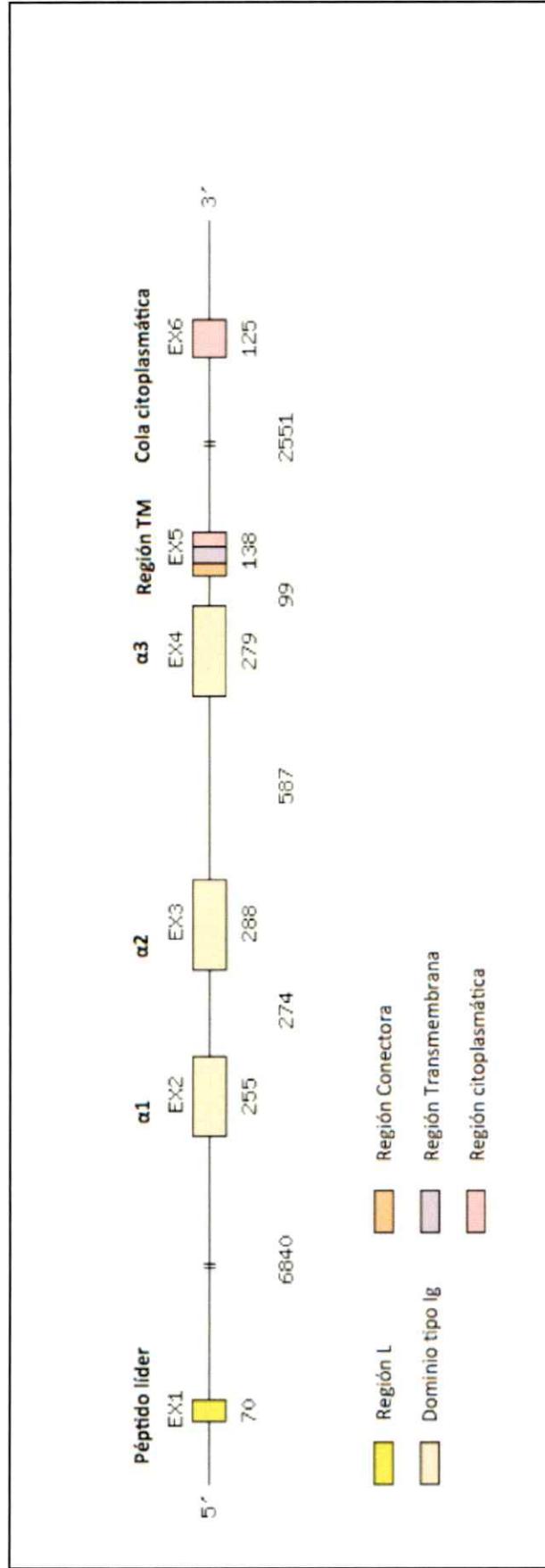


Figura 1. Organización exón-intrón del gen *MICA*. El gen *MICA* contiene 6 exones que codifican para un péptido líder (Región L, Exón 1), 3 dominios extracelulares (α 1-3) (Exones 2,3 y 4, respectivamente), un dominio de transmembrana (Exón 5) y un dominio citoplasmático (Exón 6). El largo de los exones e intrones están representados en pares de bases. (Figura modificada desde Frigoul & LeFranc, 2005).

Desde un punto de vista funcional, se ha reportado que el gen *MICA* es altamente polimórfico, documentándose hasta la fecha 91 alelos para este gen (base de datos IMGT, <http://www.imgt.org>). Se sabe que la mayoría de los sitios polimórficos varían en posiciones nucleotídicas no redundantes, con predominancia en los dominios $\alpha 2$ y $\alpha 3$ implicados tanto en la interacción con NKG2D (Figura 1) así como también con el sitio de clivaje por metaloproteasas (Salih y col., 2002). Se ha reportado que *MICA* posee variaciones en el número de repeticiones de la secuencia en tándem GCT en el exón correspondiente al segmento transmembrana y que tales variaciones o mutaciones codifican para diferente número de alaninas (Mizuki y col., 1997), dichas variaciones serían propias del desarrollo tumoral (mutaciones somáticas). En particular, el alelo A5.1 posee una inserción (G) que genera un corrimiento del marco de lectura y un codón de terminación prematuro dentro de la región transmembrana, por tal razón la proteína resultante no se expresa en la superficie de la célula, codificando una forma soluble del ligando MICA. Además, pacientes con carcinoma hepatocelular con el genotipo homocigoto A5.1 han reportado niveles altos de MICA solubles y una menor tasa de supervivencia (Pérez-Rodríguez y col., 2000; Kumar y col., 2012). Adicionalmente, el alelo *MICA**010 codifica para una proteína inestable que es degradada rápidamente dentro de la célula y que no alcanza a expresarse en la superficie celular. Se han identificado diversas sustituciones en los dominios $\alpha 1$, $\alpha 2$ y $\alpha 3$ de MICA, cuyos polimorfismos estarían implicados en la interacción con NKG2D y en particular el alelo *MICA**A9 que codifica para el SNP *MICA*-129val se ha asociado a una disminución en la afinidad a NKG2D en pacientes con Carcinoma Nasofaríngeo en la población Tunecina (Choy & Phipps, 2010).

En base a datos previos obtenidos en el laboratorio, se determinó que cerca del 70% de las células tumorales de pacientes con cáncer gástrico expresan MICA, MICB, ULBP 2-5-6, ULBP1 y ULBP3. En tanto, en el suero de pacientes con este cáncer se ha reportado la presencia de sMICA y en el sobrenadante de cultivos celulares tumorales de cáncer gástrico se ha observado la presencia tanto de MICA como MICB.

En base a estos antecedentes, el gen *MICA* ha sido caracterizado mediante secuenciación masiva de nueva generación; este proceso a diferencia de los sistemas de secuenciación tradicionales, utiliza plataformas que son capaces de generar paralelamente y de forma masiva, millones de fragmentos de ADN en un único proceso de secuenciación en un tiempo récord y por costes cada vez más reducidos. Por tal razón, y junto a su gran rendimiento, este tipo de plataformas es idóneo para un sin fin de estudios a gran escala imposibles de abordar con ningún otro tipo de tecnología existente hasta la fecha, debido al enorme coste que ello supondría. En este contexto, dos estudios de asociación del genoma completo (GWAS) han identificado un *loci* en *MICA* que estarían relacionados con susceptibilidad a neoplasia cervical y carcinoma hepatocelular inducida por el virus de la hepatitis C, respectivamente (Chen y col., 2014).

Más allá de los estudios genéticos que permiten el hallazgo de genes únicos responsables de enfermedades que siguen un patrón de herencia mendeliano, la revolución científica que ha supuesto la secuenciación masiva ha permitido realizar estudios de detección de variación propias del desarrollo de enfermedades en el individuo y en particular en el campo de la oncología, ha supuesto una gran herramienta que permite detectar



variaciones generadas durante la progresión tumoral (mutaciones somáticas) y que podrían estar dando cuenta del mecanismo por el cual el tumor se desarrolla.

Es ampliamente conocido el alto costo biológico, económico y emocional que implica la enfermedad de cáncer para el paciente y su familia, siendo a partir de los 40 años una de las principales causas de muerte en ambos sexos (superando una tasa de 2000 por cada 100.000 habitantes a los 80 años) (Medina & Kaempffer, 2000). De todos los tipos de cáncer existentes, el Cáncer Gástrico ubicó a Chile en el tercer lugar a nivel mundial en 1998 (Itriago y col, 1998), teniendo una tasa de morbilidad local de 23,2 por cada 100.000 habitantes; y en la actualidad para nuestro país, diferenciando por género, corresponde al primero en el caso de la población masculina y el quinto en la femenina (con una tasa de mortalidad de 20 por cada 100.000 habitantes) (MINSAL, 2010). El propósito del presente estudio de secuenciación masiva, es la identificación de mutaciones genéticas de tipo SNV en regiones codificantes del gen *MICA*, los que podrían estar dando cuenta de mecanismos de evasión inmune propios del desarrollo tumoral durante la progresión del adenocarcinoma gástrico; mediante una interacción alterada de NKG2D y su ligando MICA en los pacientes en estudio. Finalmente, se analizará el efecto *in silico* de estas variantes en la conformación del ligando MICA.

Los resultados de esta tesis nos ayudarán a entender los mecanismos de evasión del sistema inmune por parte del tumor y las implicaciones que en un futuro podría tener la secuenciación masiva en el diagnóstico y prevención del Cáncer Gástrico.

1.7 Hipótesis

Células tumorales de adenocarcinoma gástrico humano presentan mutaciones somáticas en el gen que codifica la proteína MICA.

1.8 Objetivo General

Identificar mutaciones somáticas en el gen de MICA en células de adenocarcinoma gástrico que expliquen la evasión inmune en pacientes con cáncer gástrico.

1.9 Objetivos Específicos

1. Secuenciar masivamente exones del gen de MICA de pacientes con cáncer gástrico.
2. Analizar la presencia de variaciones de un sólo nucleótido (SNV) del gen *MICA*.
3. Estudiar la frecuencia de mutaciones somáticas en el gen *MICA* y evaluar bioinformáticamente su efecto en la función del proteína MICA.

2. MATERIALES Y MÉTODOS

2.1. Pacientes

Para el presente estudio se reclutaron 50 pacientes diagnosticados con adenocarcinoma gástrico y sometidos a gastrectomía en el Hospital Salvador en colaboración con el Dr. Marco Bustamante, Director del Departamento de Cirugía y el Dr. Patricio González, nos proveyó de las muestras. Previo a la recolección de muestras se obtuvo una autorización firmada por el paciente, a través de un Consentimiento Informado (Anexo 1). Para el estudio, se incluyeron pacientes con tumores localizados; los que se recolectaron durante la cirugía. El criterio de exclusión de los pacientes fue haber tenido un tratamiento de quimioterapia y/o radioterapia previo a la intervención. El protocolo de estudio se aprobó por el Comité de Ética de Investigación en Seres Humanos de la Facultad de Medicina, Universidad de Chile (Anexo 2). La Tabla 1 indica los datos clínico-patológicos de los pacientes reclutados en el presente estudio.

Tabla 1. Características clínico-patológicas de los pacientes con cáncer gástrico.

		(N=50)
Edad (años)		65.66 ± 10.36
Sexo	Mujeres	16
	Hombres	34
Localización Tumor	ANTRO/ANTRAL	10
	CARDIA-ESÓFAGO	1
	CARDIAL	2
	CUERPO	4
	CUERPO Y ANTRO	3
	CURVATURA MAYOR	3
	CURVATURA MENOR	19
	FONDO	1
	FONDO-CUERPO	1
	FONDO-CUERPO-ANTRO	1
	SUBCARDIAL	3
	NA	2
	Tamaño Tumor (diámetro mayor en cm)	> o = a 5 cm
> 5 y < 10		21
< o = a 10 cm		11
Clasificación TNM	T1	6
	T2	4
	T3	12
	T4	28
<i>H. pylori</i>	+	22
	-	28

2.2. Análisis de calidad del material genético extraído desde muestras

2.2.1. Preparación del tejido tumoral

Durante la gastrectomía se obtuvieron muestras frescas de tejido tumoral de aproximadamente 3 mm³, las cuales se transportaron al laboratorio en 1 ml de amortiguador fosfato salino (PBS) pH 7,2 (Gibco, EE.UU) y se mantuvieron a -20°C hasta su uso.

2.2.2. Extracción del ADN genómico

El proceso de extracción de ADN genómico se realizó mediante el uso del kit comercial de purificación QIAamp DNA Mini Kit (QIAGEN, EE.UU) según las indicaciones del fabricante.

El protocolo inicia con la lisis de las células presentes en el tejido, mediante la incubación del volumen de muestra suspendida con Proteinasa K (20 mg/ml) en PBS a 56°C por 3 horas. El lisado celular resultante es transferido junto a 200 µl de etanol a una columna de centrifugación QIAamp Mini, con la finalidad de que el ADN genómico se una a la membrana de gel de sílice que está contenido en dicha columna. Luego, la membrana es lavada sistemáticamente con las soluciones AW1 y AW2. Finalmente se eluyó el ADN genómico depositado en la membrana con buffer AE.

2.2.3. Cuantificación de ADN mediante densidad óptica

La concentración del ADN genómico para cada muestra se determinó a través de espectroscopia UV, midiendo la absorbancia a 260 nm. Se evaluó la pureza del ADN, midiendo además la absorbancia a 280 nm, para establecer la relación A₂₆₀/A₂₈₀. Se consideraron como valores óptimos aquellos entre 1.8 y 2.0. Las mediciones se

realizaron en el equipo CYTATION 3 (Biotek Instruments, EE.UU) utilizando en cada caso 2 μ l de muestra. Como blanco de la medición se utilizó el buffer de elución AE (QIAGEN, EEUU).

2.2.4. Cuantificación de ADN de doble hebra

La medición específica de la cantidad de ADN de doble hebra contenido por muestra se realizó con el reactivo PicoGreen (Invitrogen, EE.UU). Las muestras se excitaron a 480 nm y emiten fluorescencia a 520 nm cuya intensidad se midió en el equipo CITATION 3 (Biotek Instruments, EE.UU). Este ensayo se llevó a cabo en placas negras de 96 pocillos de fondo plano (Nunc, EE.UU).

Preparación Buffer

Se preparó Buffer TE 20X (200 mM TRIS-HCl; 20 mM EDTA; pH 7.5) con agua destilada y libre de DNAsas. Luego se diluyó a una concentración final de 1X. El Buffer TE 1X (10 mM TRIS-HCl; 1 mM EDTA; pH 7.5) se utilizó para diluir el reactivo PicoGreen, las muestras de ADN y para el ensayo mismo.

Curva estándar de ADN

Como ADN estándar se utilizó el reactivo Lambda (500 μ g/ml, Invitrogen, EE.UU). En primer lugar este se diluyó a una concentración final de 5000 η g/ml en buffer TE 1X y luego se preparó la curva estándar siguiendo el protocolo que se detalla en la Tabla 2.

Tabla 2. Protocolo preparación curva estándar para ensayo PicoGreen.

Concentración η g/uL	μ L Lambda	μ L Buffer TE 1X
1000	40	60
800	32	68
600	24	76
400	16	84
200	8	92
100	4	96
50	2	98
0	0	100

Reactivo PicoGreen

El reactivo PicoGreen (2000X, Invitrogen, EE.UU) se diluyó 200 veces, el mismo día del ensayo, con buffer TE 1X preparando cantidad suficiente para agregar 100 μ l de la solución en cada pocillo. La solución se protegió de la luz ya que el reactivo PicoGreen es fotosensible.

Una vez preparada la curva, se procedió a preparar los pocillos con las muestras. Teniendo en cuenta la concentración determinada previamente mediante espectroscopía UV, se agregó 1 μ l de muestra si la concentración fue mayor a 100 ng/ μ l ó 2 μ l de muestra si la concentración fue menor a 100 ng/ μ l. Las muestras se diluyeron agregando 99 o 98 μ l de buffer TE 1X respectivamente. Finalmente, se agregó a cada uno de los pocillos 100 μ l de la solución de PicoGreen. Se incubó 5 min a temperatura ambiente y se procedió a medir la intensidad de la fluorescencia en el espectrofluorímetro. Las muestras fueron excitadas a 485 nm y la intensidad de fluorescencia fue medida a 520 nm.

2.2.5. Electroforesis de ADN en geles de agarosa

La integridad del ADN genómico extraído se evaluó mediante electroforesis en un gel de agarosa al 0,75% en TBE (Tris, borato EDTA) 1X, utilizando bromuro de etidio (0.5 μ g/ml) para la detección del ADN. En cada pocillo del gel se colocaron 2 μ l de buffer de carga con 5 μ l de ADN. En uno de los pocillos del gel se colocó como control de la concentración del ADN, 5 μ l del fago lambda (Promega EE.UU 200 μ g/ml) sin digerir. La corrida de electroforesis se realizó aplicando una corriente de 40 v por 20 min seguido de 70 v por 30 min.

2.2.6. Electroforesis Capilar de ADN

Adicionalmente se evaluó la integridad y tamaño de los fragmentos del ADN presentes en cada muestra mediante electroforesis capilar paralela. Para ello se utilizó el kit comercial High Sensitive Genomic DNA (Advanced Analytical, EE.UU) y se diluyeron las muestras de ADN genómico a una concentración no mayor a los 10 ng/μl. Siguiendo las instrucciones del fabricante las muestras diluidas se analizaron en el bioanalizador Fragment Analyzer TM (Advanced Analyticals, EE.UU), que consiste en un sistema de microelectroforesis mediante el uso de nanocapilares que separa los componentes de la muestra mediante electroforesis. Tras la corrida, el tamaño de los fragmentos de ADN se visualizaron y analizaron empleando el software PROsize 2.0 (Advanced Analyticals, EE.UU)

2.3. Secuenciación masiva de exones del gen *MICA*

2.3.1. Diseño de amplicones

Para la secuenciación dirigida de los exones 2, 3, 4 y 5 del gen *MICA*, se utilizó un kit TruSeq Custom Amplicon (Illumina Inc, EE.UU). Para ello las regiones objetivo se cubrieron por 8 amplicones, los que se ingresaron mediante coordenadas cromosómicas en el software disponible online DesignStudio (Illumina Inc, EE.UU) y posteriormente se envió una orden para el diseño del kit personalizado. Cabe señalar que al diseñar el proyecto, se agregaron 15 pares de bases en los extremos (río arriba y río abajo) de los exones de modo de cubrir completamente las regiones en estudio.

Una vez recibido el kit TruSeq Custom Amplicon (Illumina Inc, EE.UU) para el gen *MICA* se revisó el archivo manifiesto, el cual especifica el genoma de referencia y los

amplicones que cubren el gen *MICA*. Las regiones en estudio para el gen *MICA* se cubrieron con 12 amplicones, de modo que el kit de secuenciación contenía los pares de oligonucleótidos personalizados que rodean cada amplicón de la región de interés.

2.3.2. Preparación de bibliotecas

La cantidad inicial de ADN genómico utilizado por cada muestra fue de 250 ng en 10 µl de solución. Con esa cantidad de ADN (recomendada por Illumina) se prepararon las bibliotecas de las 50 muestras incluidas en este estudio. Para ello, se utilizó el kit Truseq Custom Amplicon v1.5 (Illumina Inc, EE.UU) siguiendo las indicaciones proporcionadas por Illumina.

El procedimiento de preparación de las bibliotecas se realizó de forma particular para cada muestra y se centró en las regiones genómicas de interés para el gen *MICA*. Este proceso consta de cuatro pasos clave: 1) hibridación, 2) extensión-ligadura, 3) amplificación y 4) normalización.

Durante la Hibridación, un pool de oligos específicos para el gen *MICA* hibridó (río arriba y abajo) en el ADN genómico de cada una de las muestras. Luego, empleando un filtro de selección por tamaño, se removieron por completo los oligos no unidos al ADN. A continuación el paso de extensión-ligadura consistió en enlazar los oligos que hibridaron corriente arriba y abajo de la secuencia de interés. Este proceso permitió obtener las regiones blanco flanqueadas por secuencias requeridas en las siguientes etapas.

La etapa de Amplificación consistió en amplificar el producto del paso anterior, usando partidores que permitieron agregar, en cada extremo de la hebra generada, una secuencia

índice que será utilizada para el multiplexado de muestras (i5 e i7), así como adaptadores universales requeridos para la etapa de generación de grupos o clusters (P5 y P7). Luego se purificaron los productos de PCR (denominados bibliotecas) mediante beads magnéticos, dejando fuera otros componentes de la reacción.

A continuación se procedió con el paso de Normalización, donde se regula la cantidad de cada biblioteca para garantizar una representación más equitativa de todas las muestras en la biblioteca agrupada (pool de 50 bibliotecas). Finalmente, las bibliotecas ya normalizadas se agruparon combinando volúmenes iguales de cada biblioteca. Para su secuenciación, el pool de bibliotecas agrupadas previamente se diluyeron en un buffer de hibridación (HT1) y se denaturó por calor a 96°C durante 2 minutos.

2.3.3. Secuenciación en equipo MiSeq

Para la secuenciación en el equipo MiSeq, se utilizó el kit MiSeq Reagent kit v.2 de 500 ciclos (Illumina Inc, EE.UU). Este kit incluyó el cartucho con los reactivos necesarios para la secuenciación y correspondió al lugar físico donde se cargaron las bibliotecas para ser secuenciadas. Al momento de cargar el cartucho en el secuenciador, también se cargaron los archivos relevantes para el análisis posterior a la secuenciación. Estos son: el archivo manifiesto específico de este proyecto (el que define el genoma de referencia utilizado, así como las coordenadas genómicas de cada amplicón para el gen en estudio) y el archivo hoja de muestras que especifica la combinación de los índices para cada muestra.

Las bibliotecas de simple hebra se cargaron en el pocillo 17 del cartucho de secuenciación. Las bibliotecas fluyeron a través de la celda de flujo donde hibridaron

con los oligonucleótidos anti-sentido de P5 y P7 que son complementarios a los adaptadores presentes en la secuencia de las bibliotecas.

La corrida de secuenciación se basó en la química de secuenciación por síntesis (SBS), la que utiliza un método basado en terminadores reversibles para detectar bases de nucleótidos únicos en la incorporación a las cadenas de ADN en crecimiento. Durante cada ciclo de secuenciación, se añade un desoxinucleótido trifosfato (dNTP) marcado con fluorescencia a la cadena de ácido nucleico. El nucleótido marcado sirve como terminador para la polimerización, de modo que tras cada incorporación de dNTP, el colorante fluorescente se digitaliza para identificar la base añadida (proceso denominado Llamado de Base) y a continuación una segmentación enzimática libera el terminador reversible permitiendo la incorporación del nucleótido siguiente en la secuencia. Puesto que los cuatro dNTP ligados al terminador reversible (A, G, T, C) están presentes como moléculas únicas, la competencia natural minimiza la tendencia a la incorporación. Las llamadas de bases se realizaron directamente a partir de las mediciones de intensidad de la señal durante cada ciclo de secuenciación y el resultado final es la secuenciación base por base (Figura 2).

Cabe señalar que existen diversos filtros de calidad para las bases añadidas a la secuencia, entre ellos destaca la castidad de la llamada de bases, la que consiste en la proporción de la intensidad de la señal más alta dividida entre la suma de las dos señales más altas. Si existe más de una llamada de bases con un valor de castidad inferior a 0,6 en los primeros 25 ciclos, las lecturas no pasan el filtro de calidad y no son incluidas en los datos de secuenciación.

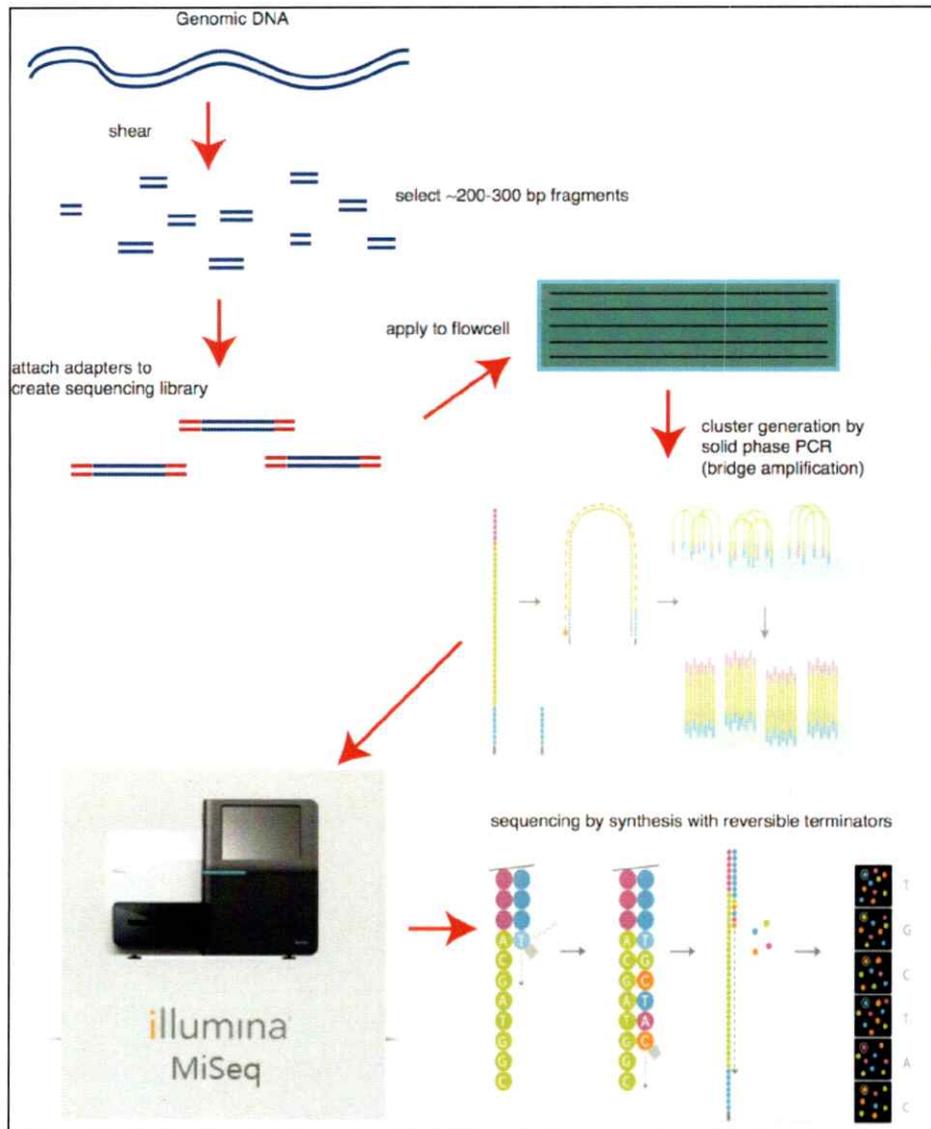


Figura 2: Flujo de trabajo durante la secuenciación masiva en el equipo MiSeq. El esquema muestra los pasos necesarios para realizar una corrida de secuenciación utilizando la tecnología MiSeq de Illumina (Modificado desde Vega y col., 2015).

2.4. Análisis Bioinformático

El análisis de imágenes y llamado de bases se realizó mediante el software de análisis en tiempo real (RTA) integrado en el equipo MiSeq. Asimismo, se asignó una puntuación de calidad a cada base incorporada en cada ciclo de secuenciación. Posteriormente, se procesaron las llamadas de bases generadas durante el análisis principal y se generó información sobre cada muestra en archivos de tipo FASTQ. Este tipo de archivo presenta un formato de texto muy similar al formato fasta, donde la Q se refiere a quality (calidad) e implica que cada posición, además de tener una letra referente al nucleótido añadido, tiene asociado un valor de calidad.

2.4.1. Comprobación de la calidad de los datos.

Las lecturas de la secuenciación se extrajeron desde el equipo MiSeq en archivos con formato FASTQ para cada muestra (Figura 3). En este tipo de archivos, la calidad de las lecturas se expresa en base logarítmica. Así, una calidad Q20 significa una posibilidad de 0.01 de haber leído una base erróneamente. Q30 traduciría una posibilidad de error de 0.001 y así sucesivamente.

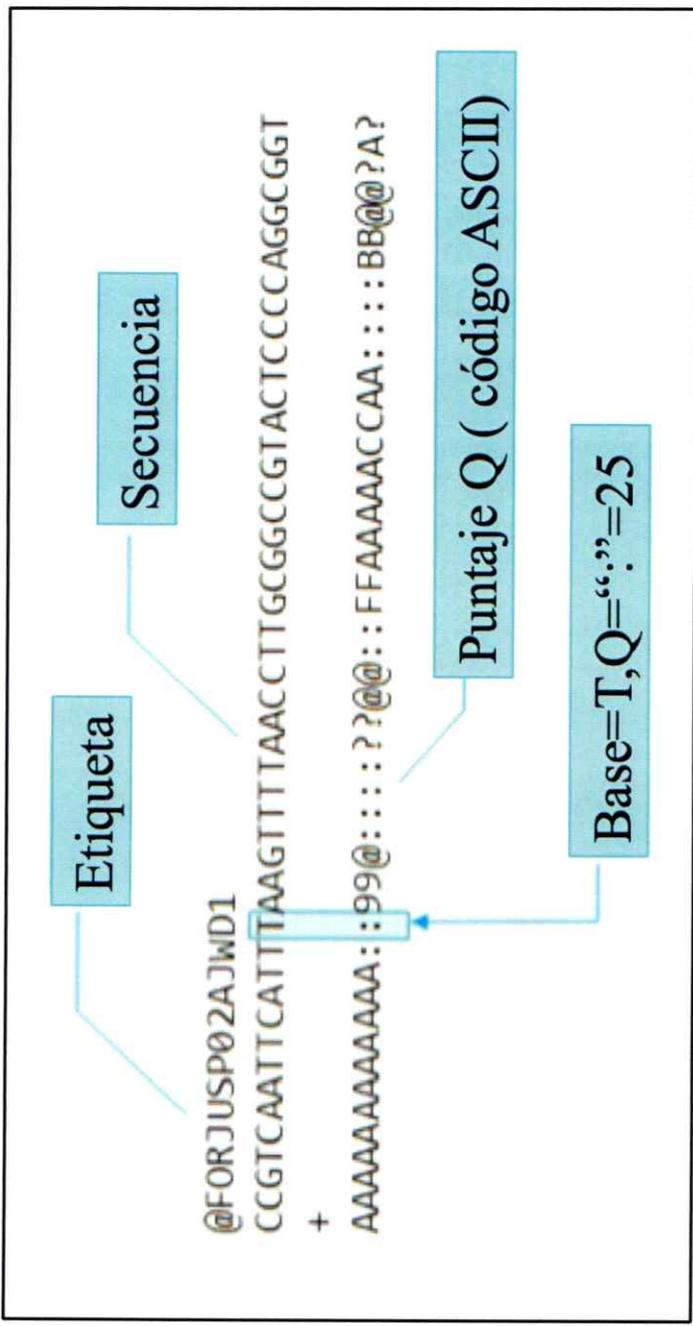


Figura 3. Representación de una lectura del secuenciador en formato FASTQ. La primera línea muestra tras la @ diversa información (Etiqueta) acerca de la máquina, el ciclo o la posición de la lectura. La segunda línea es la secuencia leída por el secuenciador. La última línea informa de la calidad de la lectura (puntaje Q) en formato ASCII (indicando la probabilidad de haber realizado una lectura errónea por cada base).

Los parámetros de calidad de los datos se exploraron con el programa FastQC (www.bioinformatics.babraham.ac.uk/projects/), que informa de los siguientes parámetros:

Calidad de la secuencia por base. Gráfica que muestra la calidad de las bases según la posición de lectura. Los datos son de buena calidad si los valores de las distintas posiciones de las lecturas son altos y se encuentra en la región verde de la grafica (en la Figura 4a las barras amarillas que muestran el rango 25-75%, son pequeñas y de tamaño semejante en todas las posiciones de lectura y se encuentran en la zona de alta puntuación). Son malos datos si la calidad de las lecturas largas es mala, observándose un aumento de variabilidad en las lecturas de las últimas posiciones de la secuencia.

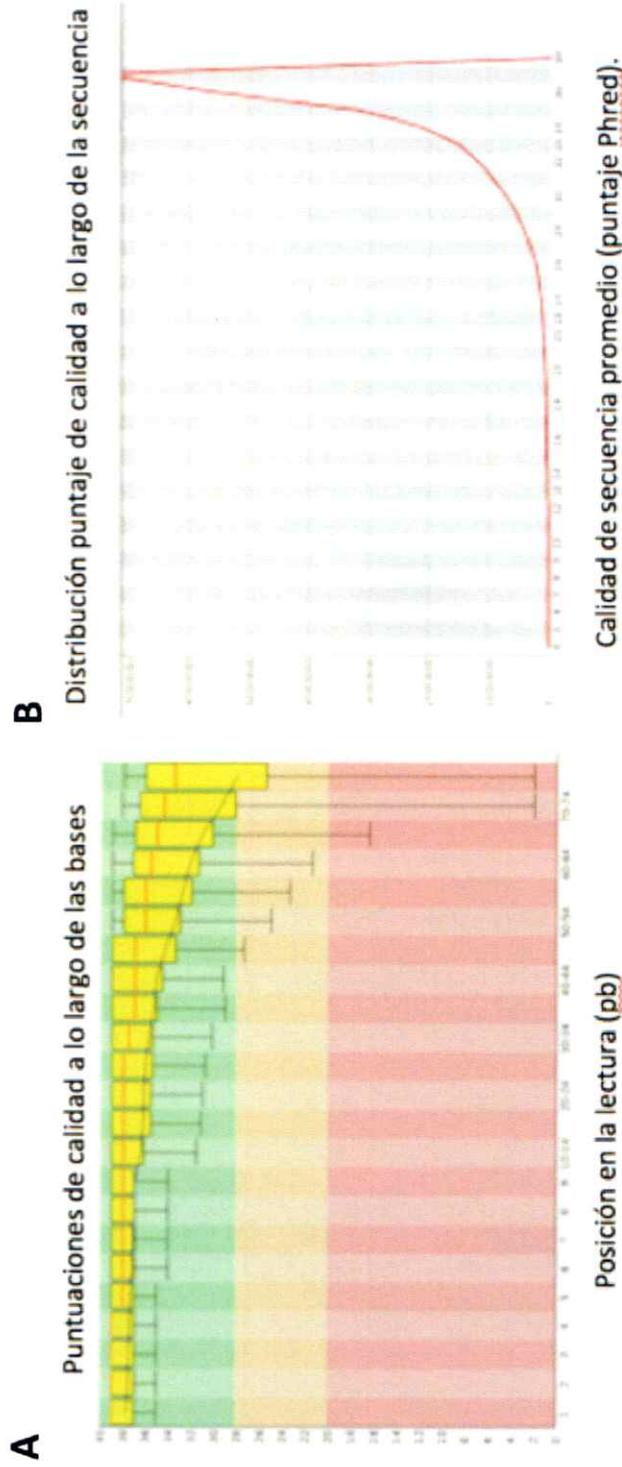


Figura 4. Gráficos de parámetros de calidad de los datos generados con el programa FastQC. **A.** Calidad de la secuencia por base. El eje x indica la posición en la secuencia en pares de bases, mientras que el eje y representa la calidad de la base para esa posición. El área roja representa bases de baja calidad, el área anaranjada representa bases con una calidad razonable, mientras que el área verde representa bases de muy buena calidad. El gráfico presentado es un ejemplo de bases con buena calidad. **B.** Distribución de la calidad de las lecturas. Este gráfico muestra la distribución de los puntajes de calidad a través de la secuencia. Incluso buenos datos de secuenciación presentan una pequeña proporción de secuencias que puede tener puntajes de baja calidad, debido a problemas propios de la química de secuenciación utilizada.

Puntuaciones de calidad por secuencia. Gráfica que representa la distribución de la calidad de las lecturas (Figura 4b). Si una gran cantidad de secuencias en un ciclo determinado son de baja calidad, puede deberse a un problema sistemático del proceso de secuenciación.

2.4.2. Alineación de las lecturas con el genoma de referencia.

Para el análisis de datos, se extrajeron los archivos de secuencia FastQ desde el equipo MiSeq y se subieron al servidor BaseSpace provisto por Illumina (<https://basespace.illumina.com/home/index>), el cual permite el almacenamiento y análisis de los datos de secuenciación.

En este punto se utilizó la aplicación TruSeq Amplicon App para llevar a cabo el análisis de datos de secuencia (Figura 5), de acuerdo al flujo de trabajo integrado Amplicon TruSeq (Illumina Inc, EE.UU).

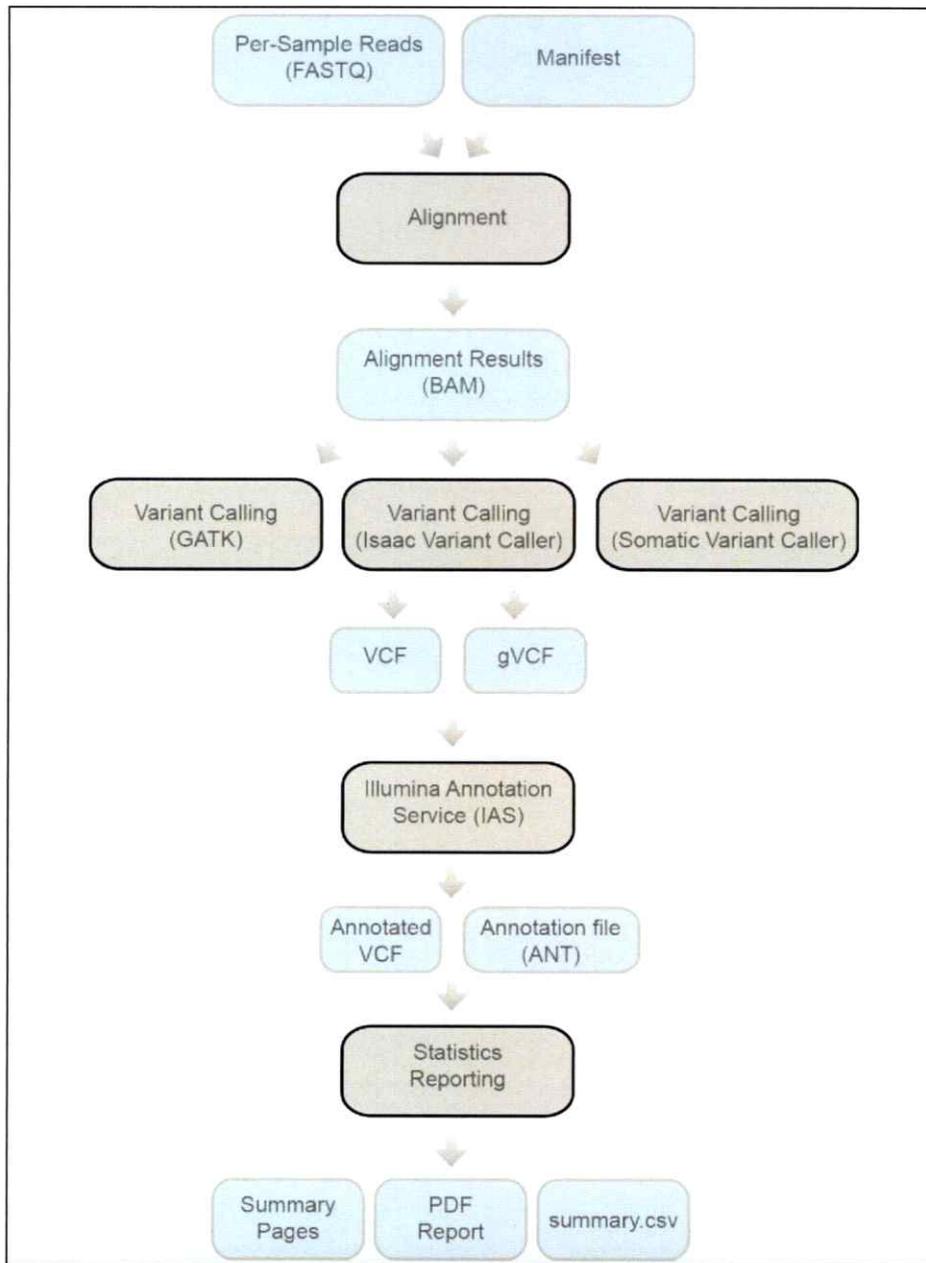


Figura 5: Herramientas informáticas utilizadas en el análisis de los datos. Los archivos de secuencias (Fastq) generados durante la secuenciación fueron alineados al genoma de referencia (Hg19). Para el llamado de variantes se utilizó el algoritmo Somatic Variant Caller (SVC) y las variantes detectadas fueron reportadas en archivos de tipo VCF (Variant call format).

La alineación con el genoma de referencia (Hg19) se realizó a lo largo de la longitud de las secuencias objetivo de amplicones mediante un algoritmo de Smith-Waterman en bandas. Esto generó un archivo BAM (*.bam) por cada muestra, el que contendrá datos de alineación de secuencias.

2.4.3 Localización de las variantes presentes en el gen *MICA*.

La identificación de todas las variaciones (llamado de variante) presentes en el gen *MICA* para cada paciente se realizó utilizando el algoritmo Somatic Variant Caller (Figura 5), recomendado para el análisis de datos generados con el kit utilizado en este proyecto (TruSeq Custom Amplicon, Illumina Inc. EE.UU). Este algoritmo identifica aquellas posiciones cromosómicas donde la secuencia generada no coincide con el genoma de referencia (Hg19) y las posiciones cromosómicas donde se detecta un cambio de base se reportan en un archivo denominado formato de llamadas de variante (VCF). El archivo VCF se obtuvo mediante el uso del Servicio de Anotación de Illumina (IAS) específicamente para las regiones incluidas en el archivo Manifiesto, el que presenta las coordenadas cromosómicas de las regiones del gen en estudio. Por este flujo de trabajo se obtuvieron documentos de tipo BAM (archivo de alineamiento), VCF (archivo que contiene las variantes detectadas) y summary.csv (archivo que contiene información de la secuenciación y variantes para cada muestra).

2.4.4. Filtrado de las variantes detectadas

Si bien los archivos VCF generados por este flujo de trabajo contenían anotaciones para variantes localizadas en las regiones del gen *MICA* en estudio, se utilizó de forma adicional el software Variant Studio 2.2 (Illumina Inc, EE.UU) para seleccionar sólo

aquellas variantes de tipo SNV que pasaron (PASS) todos los filtros descritos en la Tabla 3 y que se localizaran en el regiones exónicas del gen. El filtrado según esta serie de parámetros permitió eliminar lo que se pueden considerar falsos positivos (defectos de la técnica).

Tabla 3: Lista de filtros utilizados en la detección de variantes

Tipo de Filtro	Efecto
<i>"LowCoverage"</i> o filtro de baja cobertura.	Filtra variaciones con coberturas menores de un número a seleccionar de lecturas, ya que son potenciales artefactos. El valor por defecto es de 5 lecturas.
<i>"VeryLowQual"</i> o filtro de muy baja calidad.	Elimina aquellas variaciones con una puntuación de calidad de menos de 30, que suelen ser artefactos
<i>"LowQual"</i> o filtro de baja calidad	Elimina aquellas variaciones con puntuaciones de calidad entre 30 y 50, que pueden ser artefactos.
<i>"LowQD"</i> o filtro de baja QD (confianza de la variante/profundidad no filtrada).	Puntuaciones bajas del parámetro QD suelen representar falsos positivos. En este caso, elimina las variantes con puntuaciones QD por debajo de 1.5.
<i>"StrandBias"</i> .	Variaciones que sólo aparecen en las lecturas de la misma dirección son habitualmente artefactos, por lo que se filtran.

Se generó un archivo VCF definitivo que contiene las variantes de interés junto con información sobre cada una de ellas. El archivo con extensión .csv se puede leer en formato Excel e incluye los siguientes datos de cada variante identificada: gen, cambio de aminoácido, frecuencia en la que el alelo aparece en la base de datos de los 1000 genomas, referencia dbSNP y puntuaciones de predicción (según programas como PolyPhen), entre otros (Figura 6).



Variantes-MICA_detalle.xlsx

125% Revisar

Inicio Diseño Tablas Gráficos SmartArt Fórmulas Datos Revisar

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1	Sample	Gene	Variant	Chr	Coordinate	Type	Genotype	Exonic	Filters	Alt Variant Freq	Read Depth	Alt Read Depth	Allelic Depths	Sift	PolyPhen	dbSNP ID
2	CG01	MICA	T>C	6	31371430	snv	hom	yes	PASS	99.87	3886	3881	13.881			rs114021487, rs2844519
3	CG01	MICA	T>T/G	6	31378358	snv	het	yes	PASS	39.39	2898	1142	17.501.142	tolerated(0.07)	benign(0.26)	rs116741206, rs1063630
4	CG01	MICA	A>G	6	31378387	snv	het	yes	PASS	39.58	2903	1149	17.531.149			rs1063631, rs114438708
5	CG01	MICA	A>G	6	31378425	snv	het	yes	PASS	39.78	2921	1162	17.581.162	tolerated(0.18)	benign(0.068)	rs1051786, rs116048399
6	CG01	MICA	G>G/A	6	31378077	snv	het	yes	PASS	49.58	4508	2235	22.712.235	possibly_damaging(0.730)		rs114755437, rs1051792
7	CG01	MICA	G>G/A	6	31379109	snv	het	yes	PASS	39.58	1824	722	1.100.722	tolerated(0.55)	benign(0.05)	rs1051794, rs114163881
8	CG01	MICA	G>G/A	6	31379931	snv	het	yes	PASS	52.8	2377	1255	11.211.255	tolerated(1)	benign(0.004)	rs1063635
9	CG01	MICA	T>T/C	6	31380194	snv	het	yes	PASS	43.42	9011	3913	50.883.913	deleterious(0)	unknown(C)	rs41554412, rs114717570
10	CG01	MICA	C>C/T	6	31380198	snv	het	yes	PASS	43.16	9015	3891	51.223.891	deleterious(0)	unknown(C)	rs61738275, rs115720194
11	CG01	MICA	A>A/G	6	31382770	snv	het	yes	PASS	27.61	6534	1804	47.231.804	deleterious(0)	unknown(C)	rs116179804, rs9266824
12	CG01	MICA	C>C/A	6	31382882	snv	het	yes	PASS	27.01	7375	1952	53.691.592			rs1148771043, rs9266825
13	CG01	MICA	A>A/G	6	31382911	snv	het	yes	PASS	27.45	7366	2023	54.412.023			rs115463816, rs1882
14	CG01	MICA	G>G/A	6	31382932	snv	het	yes	PASS	27.1	7436	2015	54.182.015			rs116777003, rs9266826
15	CG01	MICA	A>A/G	6	31382938	snv	het	yes	PASS	27.32	7383	2017	53.632.017			rs9266827, rs114288468
16	CG01	MICA	G>G/C	6	31382942	snv	het	yes	PASS	27.03	7469	2019	54.362.019			rs116398930, rs9266828
17	CG01	MICA	A>A/G	6	31382964	snv	het	yes	PASS	27.13	7475	2028	54.312.028			rs9266829, rs115831270
18	CG02	MICA	T>C/C	6	31371430	snv	hom	yes	PASS	100	1124	1124	0.1124			rs114021487, rs2844519
19	CG02	MICA	G>G/A	6	31378301	snv	het	yes	PASS	53.65	1521	816	705.816			rs145823885, rs2853992
20	CG02	MICA	T>T/G	6	31378358	snv	het	yes	PASS	45.73	1640	750	885.75	tolerated(0.07)	benign(0.26)	rs116741206, rs1063630
21	CG02	MICA	T>T/C	6	31378387	snv	het	yes	PASS	46.1	1653	762	891.762			rs1063631, rs114438708
22	CG02	MICA	A>A/G	6	31378425	snv	het	yes	PASS	46.33	1650	769	891.769	tolerated(0.18)	benign(0.068)	rs1051786, rs116048399
23	CG02	MICA	C>C/G	6	31378956	snv	het	yes	PASS	56.55	4596	2599	19.832.599	tolerated(0.26)	probably_damaging(0.997)	rs114304655, rs1051790
24	CG02	MICA	G>G/A	6	31378977	snv	het	yes	PASS	43.09	4614	1988	26.261.988	deleterious(0.03)	possibly_damaging(0.739)	rs114755437, rs1051792
25	CG02	MICA	G>G/A	6	31379109	snv	het	yes	PASS	28.6	1203	344	858.344	tolerated(0.55)	benign(0.05)	rs1051794, rs114163881
26	CG02	MICA	G>G/A	6	31379115	snv	het	yes	PASS	69.79	1218	850	360.85	tolerated(0.45)	benign(0.09)	rs116204553, rs1131896
27	CG02	MICA	C>C/G	6	31379134	snv	het	yes	PASS	70.45	1212	854	356.854	deleterious(0.04)	benign(0.028)	rs115257533, rs1131897
28	CG02	MICA	C>C/G	6	31379141	snv	het	yes	PASS	69.86	1231	860	370.86			rs2256026, rs114783026
29	CG02	MICA	G>A/A	6	31379931	snv	hom	yes	PASS	99.39	1482	1473	41.473	tolerated(1)	benign(0.004)	rs1063635
30	CG02	MICA	T>T/C	6	31380194	snv	het	yes	PASS	42.32	4338	1836	25.001.836	deleterious(0)	unknown(C)	rs41554412, rs114717570
31	CG02	MICA	C>C/T	6	31380198	snv	het	yes	PASS	41.85	4337	1815	25.201.815	deleterious(0)	unknown(C)	rs61738275, rs115720194
32	CG02	MICA	A>A/G	6	31382770	snv	het	yes	PASS	26.96	3553	958	2.590.958			rs116179804, rs9266824

Figura 6. Visualización de archivo VCF en Excel. En azul se destaca la columna que indica el nombre del gen estudiado, en este caso MICA. En amarillo se destacan los filtros utilizados para variantes de tipo SNV, ubicadas en regiones exónicas y que pasen los filtros descritos en la Tabla 3

2.4.5 Identificación de mutaciones somáticas.

A partir de la premisa que frecuencias alélicas distintas de una proporción mendeliana, 50/50 o 100/0, estarían dando cuenta de mutaciones somáticas. En el archivo VCF se seleccionaron aquellas variaciones que presentaron una frecuencia alélica menor al 35% (Yost y col., 2012), las cuales estarían representando mutaciones propias del tumor

2.4.6. Predicción del efecto de mutaciones somáticas en la proteína MICA

Algoritmos SIFT y PROVEAN

Con el fin de determinar el efecto de variaciones no sinónimas en la proteína MICA, se utilizó el algoritmo SIFT v1.03 (Sorting Tolerant From Intolerant) (disponible en <http://sift.jcvi.org>), el cual simula el efecto fenotípico de sustituciones aminoacídicas bajo la condición de que existe un alto grado de conservación en regiones con posiciones importantes para la proteína.

Adicionalmente, se utilizó el algoritmo PROVEAN v1.1.3 (Protein Variation Effect Analyzer) (<http://provean.jcvi.org/index.php>), el cual es un software que predice si una sustitución de aminoácido tiene un impacto en la función biológica de una proteína.

Base de datos SNPeffect

Se utilizó la base de datos SNPeffect 4.0 (<http://snpeffect.switchlab.org/>) para determinar el efecto de variaciones de tipo SNV presentes en la base de datos dbSNP sobre la proteína MICA, para ello se evaluó el impacto que tienen en la proteína con una variedad de algoritmos.

Los algoritmos que se utilizaron proporcionan información sobre una serie de características de la proteína, entre ellos, SMART y Pfam permiten ver si la variación se produce dentro de los dominios funcionales de la proteína.

Visualizador molecular PyMOL

Para determinar el efecto de las variaciones detectadas en la estructura tridimensional de la proteína se utilizó el visualizador molecular PyMOL ejecutado en el entorno de programación Python; para determinar mediante animaciones, los aspectos referentes a cambios en la estructura por variaciones no sinónimas en la cristalografía de la proteína MICA (PDB 1B3J).

2.4.7. Visualización de los resultados.

Se utilizó el programa IGV (Integrative Genomics Viewer) (Robinson y col., 2011) que muestra de forma gráfica los datos de secuenciación obtenidos (Figura 7). Además, permite diferenciar con mayor facilidad verdaderas variaciones frente a artefactos. Igualmente, facilita la comparación entre muestras.

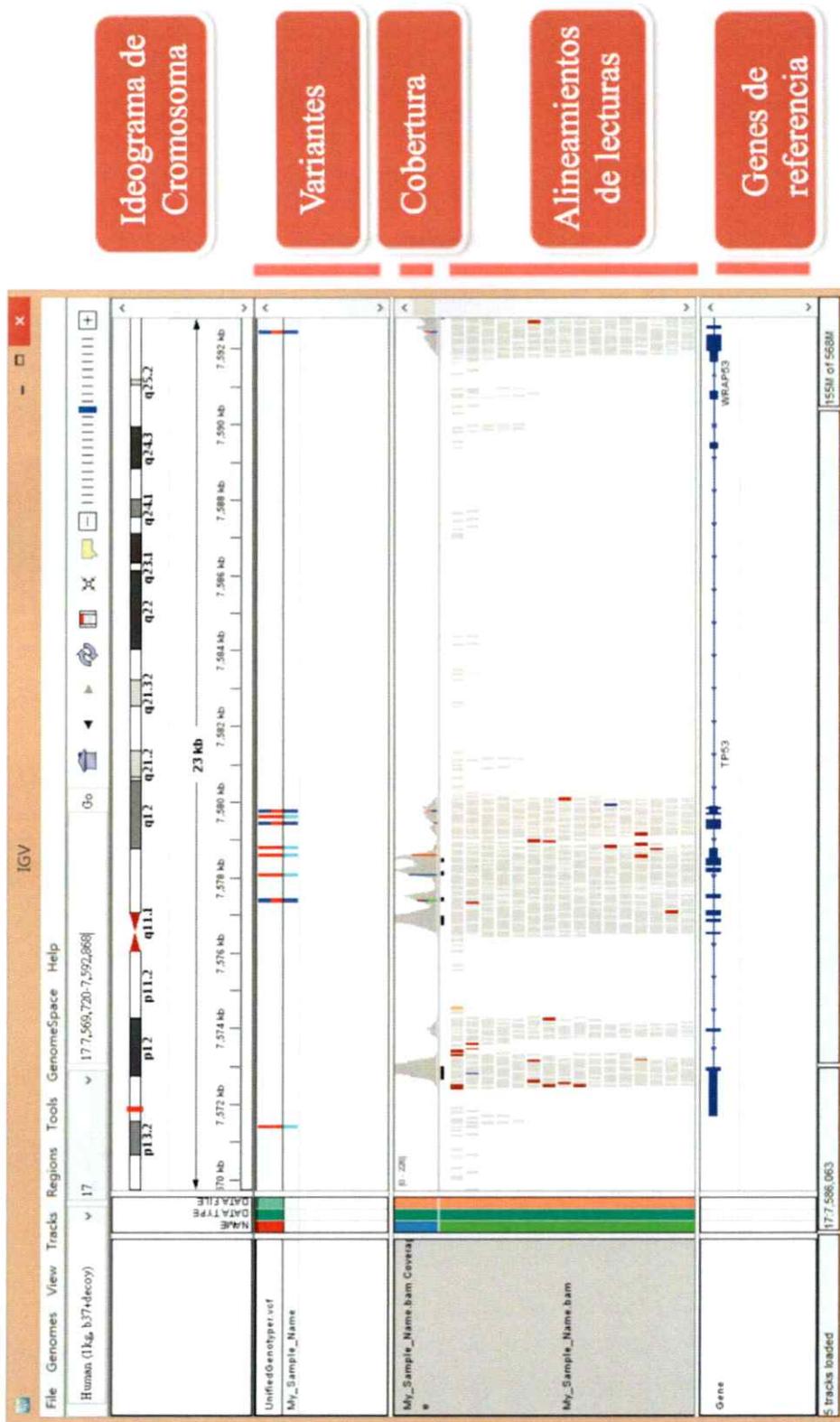


Figura 7. Ejemplo de visualización de los datos con el programa IGV. Se muestra el área cromosómica correspondiente y en la parte superior muestra un histograma de la profundidad de cobertura de las lecturas para cada posición.

3. RESULTADOS

3.1 Selección de muestras en estudio y secuenciación masiva de exones del gen *MICA*

Durante este estudio, se analizaron 50 pacientes con adenocarcinoma gástrico sometidos a gastrectomía total, 16 mujeres y 34 hombres, con un promedio de edad de 65,1 años. La Tabla 1 resume los diversos datos clínicos, patológicos y operatorios correspondientes a los 50 pacientes. Sin embargo, cabe señalar que en un comienzo se trabajó con 60 individuos a partir de los cuales se seleccionó el grupo de individuos en estudio; tal selección se realizó en base a una serie de características utilizadas como control de calidad, que daban cuenta que el ADN de la muestra era óptimo para ser secuenciado, como se describe a continuación.

3.1.1 Integridad del material genético

El primer criterio para seleccionar las muestras en estudio fue la integridad del material genético, la que se analizó mediante electroforesis.

El ADN extraído desde muestras del tejido tumoral de los pacientes con adenocarcinoma gástrico se visualizó en geles de agarosa, como se describe en la sección Materiales y Métodos. A partir de los geles obtenidos, las 60 muestras se separaron en 5 grupos de acuerdo a la integridad del ADN genómico (ADNg) obtenido de las muestras (Figura 8), en los Grupos I y II se presentó una banda bien definida de ADN genómico de alrededor de 20.000 pares de bases (pb), mientras que en el grupo V no se logró ver una banda, por tal razón las muestras de este grupo se descartaron del estudio.

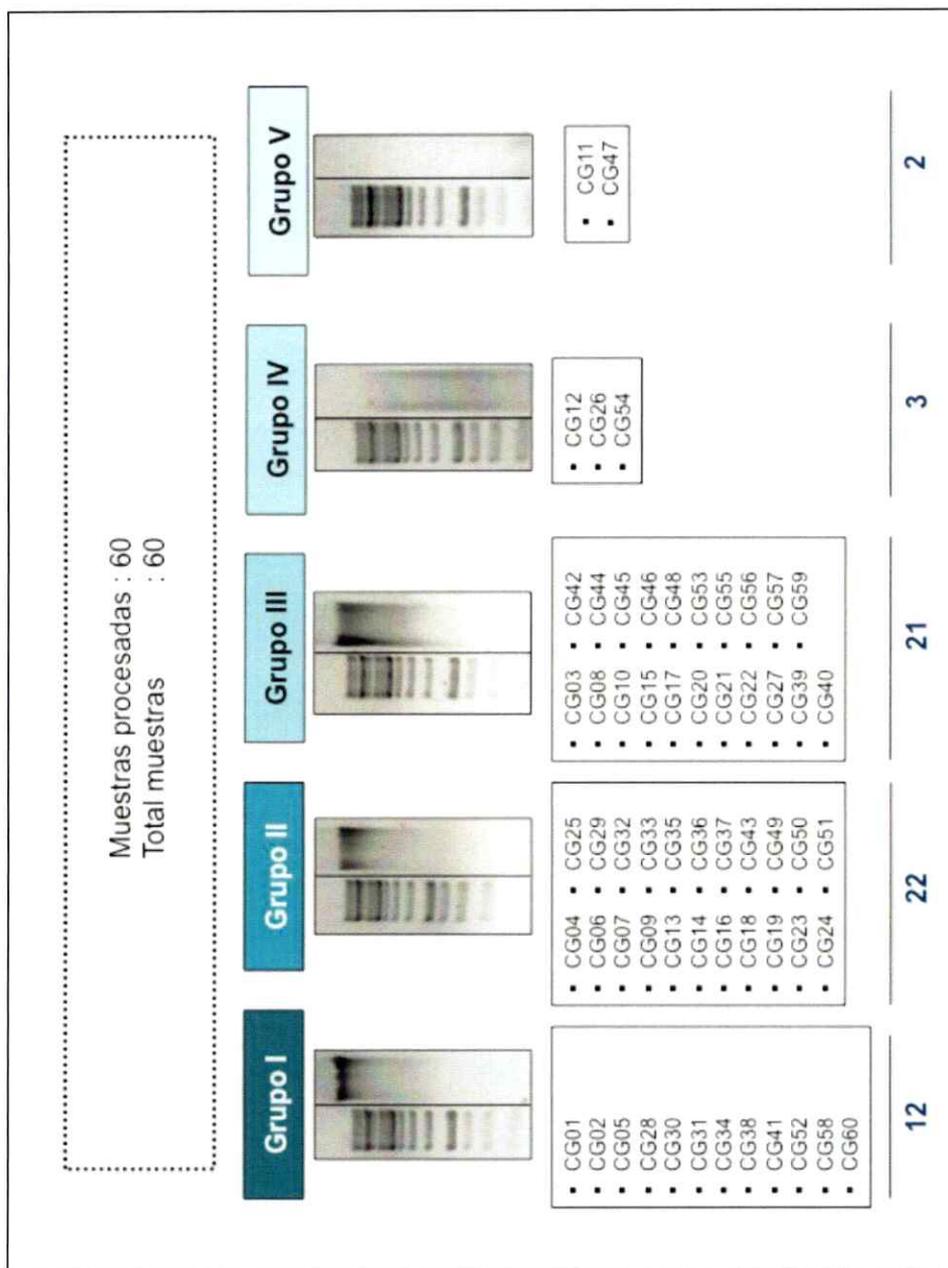


Figura 8. Agrupación de muestras en estudio de acuerdo a la integridad del material genético. Visualización de ADN genómicos, mediante electroforesis en gel de agarosa (0,5%) en buffer TBE IX, tratado con bromuro de etidio. El Grupo V se descartó del estudio debido a que no se visualizaron fragmentos de ADNg.

Por otro lado, la integridad total del ADN se evaluó mediante un electroforesis capilar. La que consiste en un sistema de microelectroforesis mediante el uso de nanocapilares, que separa los componentes de la muestra mediante electroforesis; de este modo se logró visualizar con más detalle el rango de fragmentos de ADN_g para cada muestra.

Como se muestra en la Figura 9, se observa que para los Grupos I y II los resultados se correlacionaron a lo observado en los geles de agarosa, obteniéndose un rango estrecho con fragmentos mayores a 7000 pb. En tanto, para los Grupos III y IV, se observó que el rango de fragmentos es más amplio llegando incluso a presentar tamaños cercanos a los 100 pb (Figura 10).

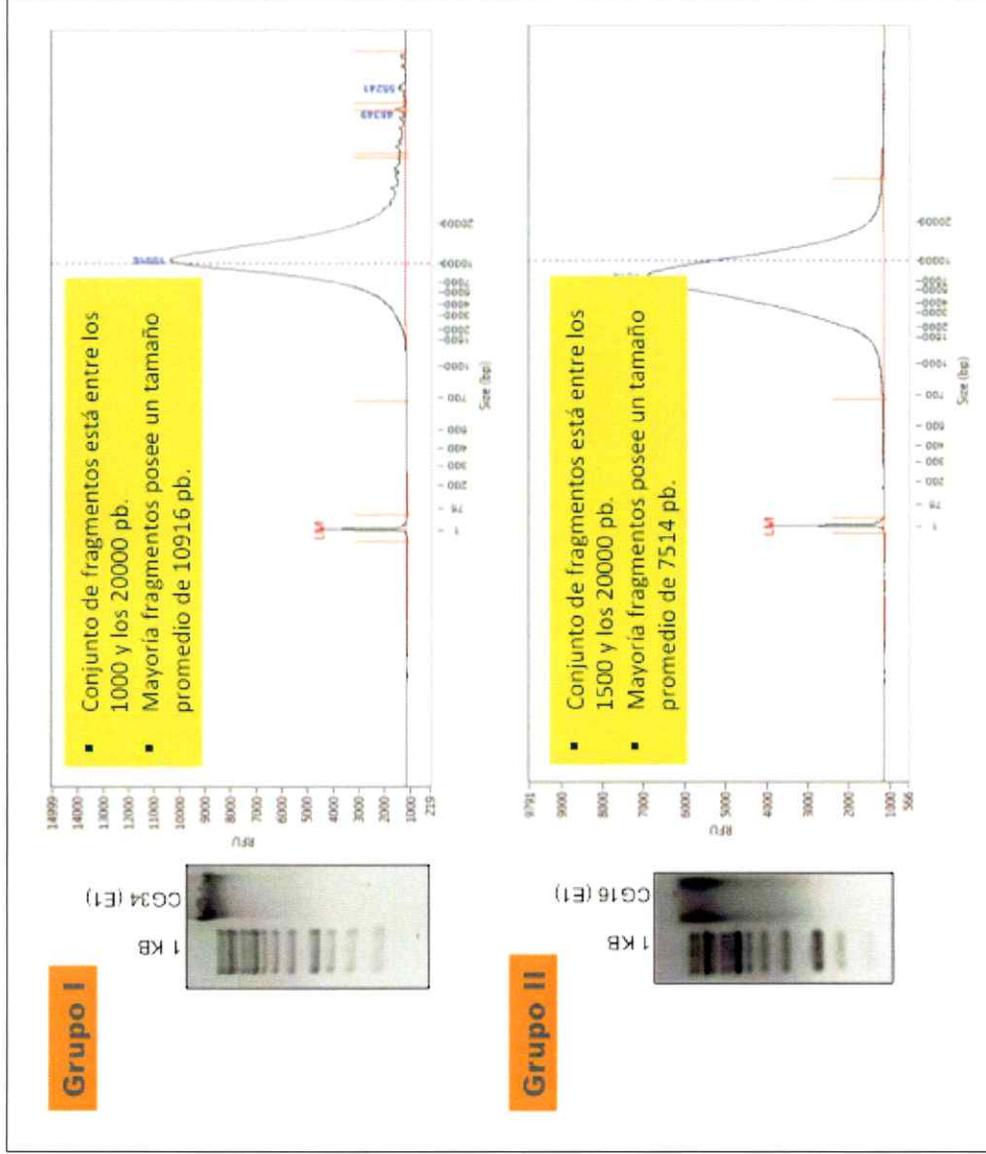


Figura 9: Distribución de fragmentos de ADN genómico en muestras del Grupo I y II. Visualización de ADN genómico mediante electroforesis capilar en un bioanalizador Fragment Analyzer (Advance Analyticals, EE.UU). Ambos Grupos presentaron fragmentos de ADN genómico con un rango de tamaño promedio mayor a 7000 pb.

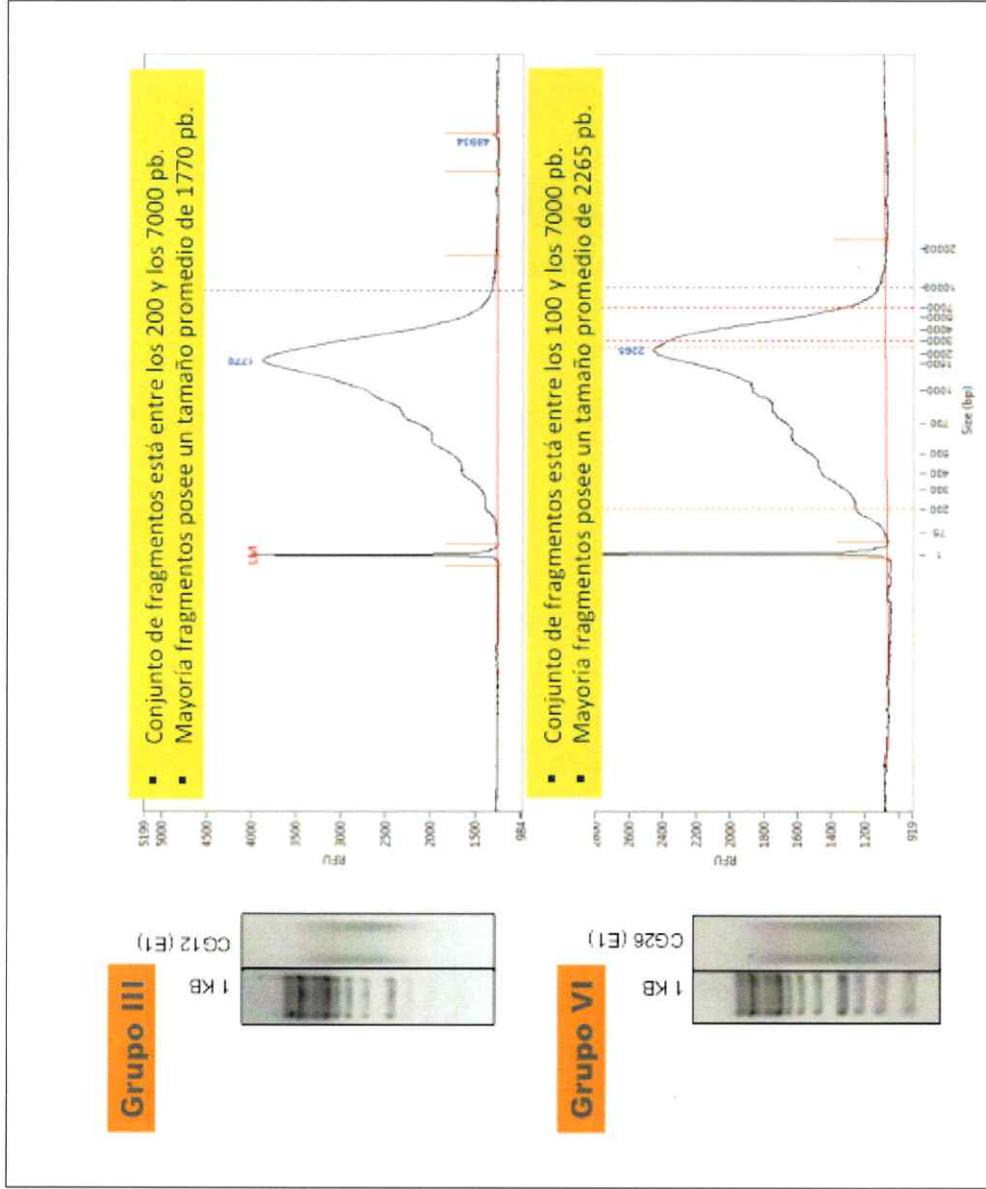


Figura 10. Distribución de fragmentos de ADN genómico en muestras del Grupo III y VI. Visualización de ADN genómico mediante electroforesis capilar en un bioanalizador Fragment Analyzer (Advance Analyticals, EE.UU). Ambos Grupos presentaron fragmentos de ADNg con un tamaño promedio mayor a 1500 pb, sin embargo, muestran un rango de tamaño mayor y se presentaron fragmentos cercanos a los 100 pb.

Adicionalmente se cuantificó el ADN de doble hebra de las muestras en estudio, midiéndose la emisión de fluorescencia a 520 nm. En base a los resultados se seleccionaron sólo aquellas muestra que presentaron una concentración mayor a 50 ng/μl, debido a que ésta es la concentración mínima que sugiere Illumina para muestras a ser secuenciadas.

En base a estos resultados se obtuvo la tabla IV, que da cuenta en detalle de las características de cada muestra analizada. Se indica el grupo al que pertenece la muestra de acuerdo a la clasificación en gel de agarosa y se presentan los resultados de la cuantificación. En rojo se destacan las muestras que presentaron concentraciones menores a 50 ng/μl, por lo que no fueron incluidas en el estudio.

Cabe señalar que muestras pertenecientes a los grupos I y III que presentaron en la electroforesis capilar fragmentos de ADN mayores a 1000 pb, al ser cuantificadas con el reactivo Picogreen que mide la concentración de ADN de doble hebra, presentaron una concentración menor a 50 ng/μl, valor mínimo requerido por Illumina para la secuenciación. Por lo que estas muestras no fueron seleccionadas para ser secuenciadas.

Tabla 4: Detalle de la cuantificación de cada muestra analizada.

CLASIFICACIÓN DE MUESTRAS POR GEL AGAROSA		CUANTIFICACION		MUESTRAS SELECCIONADAS PARA SECUENCIACIÓN	CLASIFICACIÓN DE MUESTRAS POR GEL AGAROSA		CUANTIFICACION		MUESTRAS SELECCIONADAS PARA SECUENCIACIÓN
GRUPO	MUESTRA	Espectroscopia UV 260 nm	Fluorescencia Emisión 520 nm		GRUPO	MUESTRA	Espectroscopia UV 260 nm	Fluorescencia Emisión 520 nm	
I	GG01	229.684	121.937	X	III	GG03	161.298	47.704	
	GG02	211.929	114.255	X		GG08	277.006	13.682	X
	GG05	76.683	12.872			GG10	72.095	46.356	X
	GG28	392.529	184.764	X		GG15	286.677	182.839	X
	GG30	43.650	26.896	X		GG17	329.153	190.348	X
	GG31	8.308	6.729			GG20	68.386	33.176	
	GG34	474.174	193.929	X		GG21	140.839	82.490	X
	GG38	220.195	128.608	X		GG22	310.284	162.618	X
	GG41	233.183	113.652	X		GG27	127.637	86.316	X
	GG52	411.113	139.037	X		GG39	331.461	182.687	X
	GG58	369.717	293.122	X		GG40	532.102	199.075	X
	GG60	359.383	290.705	X		GG42	132.929	70.107	X
	GG04	356.277	83.264	X		GG44	173.825	111.816	X
	GG06	448.604	194.940	X		GG45	147.027	61.292	X
GG07	279.122	151.533	X	GG46	333.355	177.473	X		
GG09	301.248	154.692	X	GG48	132.994	30.599			
GG13	221.050	136.497	X	GG53	285.874	144.787	X		
GG14	235.300	115.821	X	GG55	105.237	88.949	X		
GG16	244.583	144.921	X	GG56	201.828	164.909	X		
GG18	257.631	156.393	X	GG57	222.576	201.948	X		
GG19	436.176	194.662	X	GG59	98.705	102.534	X		
GG23	264.670	152.312	X	GG12	178.599	38.195			
GG24	229.105	113.606	X	GG26	100.648	40.282			
GG25	259.563	138.999	X	GG54	136.048	41.343			
GG29	197.523	110.422	X	GG11	19.131	12.904			
GG32	96.998	64.776	X	GG47	20.431	13.422			
GG33	157.303	93.701	X						
GG35	310.316	167.293	X						
GG36	283.695	115.101	X						
GG37	293.985	145.030	X						
GG43	221.132	116.403	X						
GG49	417.292	163.389	X						
GG50	539.347	193.398	X						
GG51	267.762	179.242	X						

TOTAL MUESTRAS PROCESADAS	60
MUESTRAS SELECCIONADAS	50

En base a los análisis presentados, de las 60 muestras procesadas, sólo 50 cumplieron con los criterios de integridad y cantida de ADN genómico que las incluye en el grupo seleccionado para ser secuenciado (Figura 11).

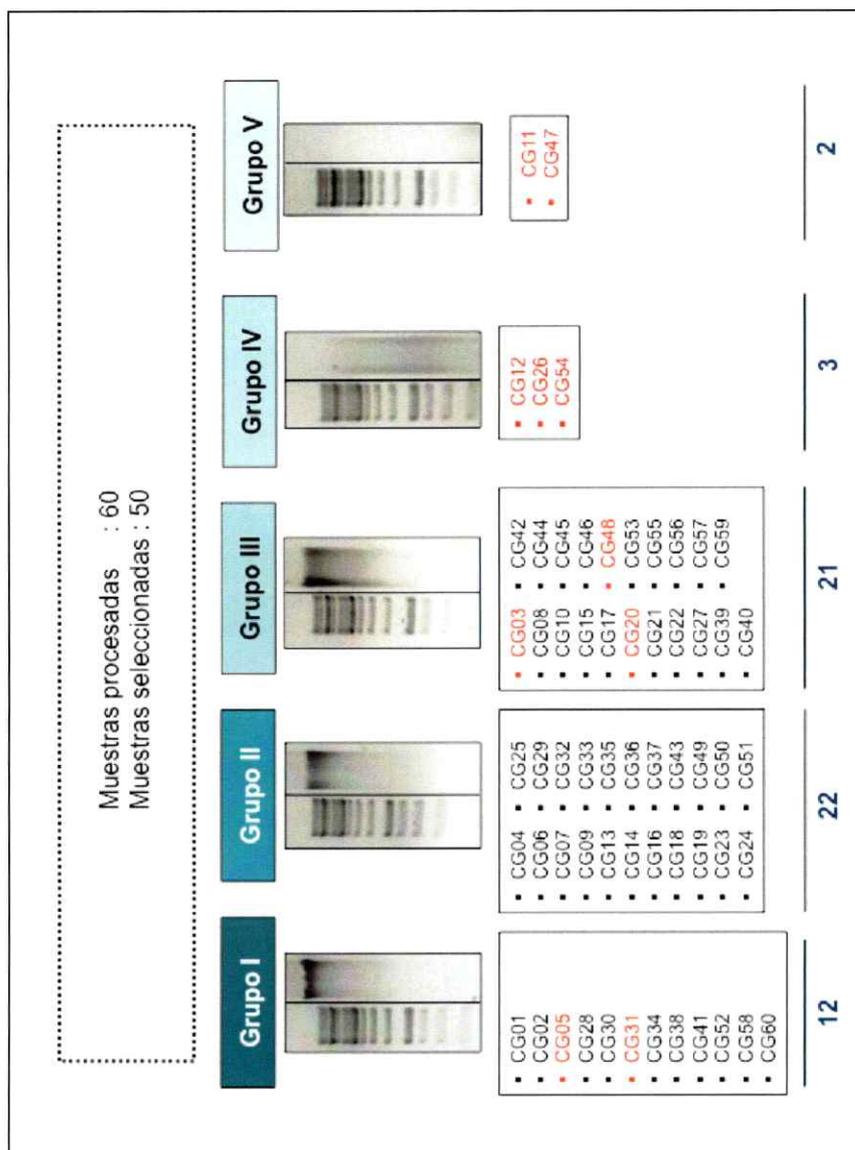


Figura 11. Selección de las muestras a ser secuenciadas. De las 60 muestras procesadas, 50 fueron seleccionadas para ser secuenciadas en base a sus perfiles de integridad de ADN genómico visualizado mediante electroforesis en gel de agarosa y electroforesis capilar. En rojo se destacan las muestras que no cumplieron con los criterios mínimos de integridad de ADNg para ser secuenciado, ni con las concentraciones de ADNg mínimas requeridas para el proceso.

3.1.2 Diseño de amplicones.

El proyecto original se diseñó con el fin de cubrir los exones 2, 3, 4, 5 del gen *MICA*. Sin embargo, al recibir los reactivos personalizados del kit TruSeq Custom Amplicon (Illumina Inc, EE.UU) junto al archivo Manifiesto que contiene las coordenadas cromosómicas del estudio, se evidenció que se adicionaron amplicones que cubrían regiones del exón 1 y del exón 6 del gen *MICA* (Figura 12).

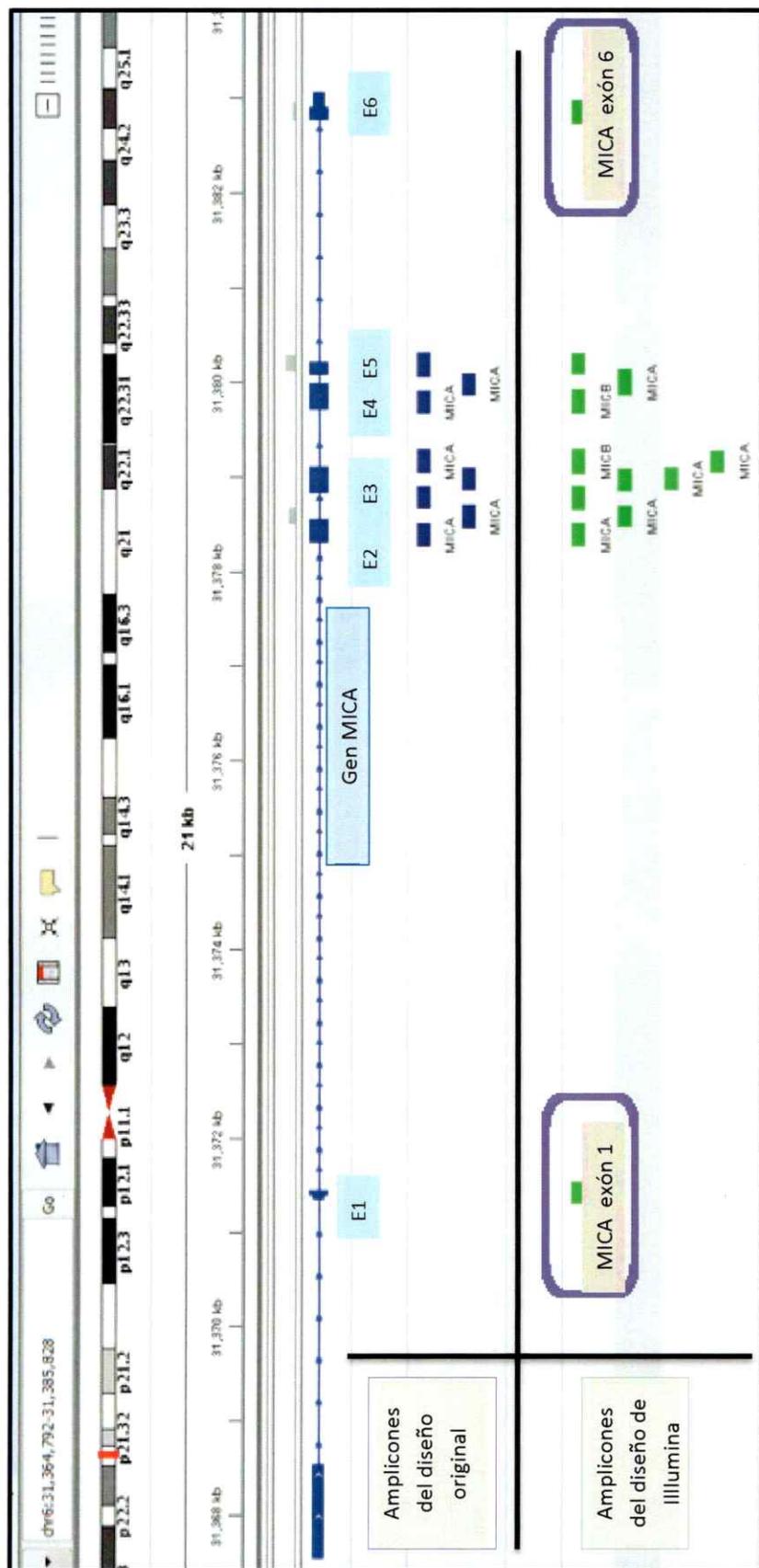


Figura 12. Visualización de regiones adicionales en el estudio del gen *MICA*. En azul se presenta el gen *MICA* y la disposición de sus exones, las bandas moradas dan cuenta de los amplicones del diseño original que cubrían los exones 2, 3, 4 y 5 del gen. Las bandas en verde representan los amplicones enviados por Illumina e incluidos en el proyecto; se destacan en un recuadro morado los nuevos amplicones que cubren los exones 1 y 6 del gen. Los amplicones se visualizaron con el programa IGV versión 2.3.

En el caso de los exones en estudio para el gen *MICA*, un análisis más detallado del total de amplicones incluidos en el proyecto dió cuenta que se adicionaron 2 nuevos amplicones que mapearon regiones de interés para el gen *MICA*, en particular en el exón 3 (Figura 13).

3.1.3 Métricas de la secuenciación

Durante la corrida de secuenciación se evaluó en tiempo real la calidad del proceso. Una vez que las bibliotecas fueron cargadas en el cartucho de secuenciación, se generaron los grupos o clusters para cada una de ellas. Por tal razón, un parámetro relevante consistió en evaluar la calidad de los grupos generados, de forma que se eliminen los datos menos fiables basados en el filtro de castidad de Illumina, el que elimina las lecturas que no estén en fase con el resto de las lecturas del cluster (Figura 14). Por tal razón del total de lecturas generadas, sólo aquellas que pasen el filtro de castidad se considerarán como lecturas passing filter (PF) y se considerarán para el futuro análisis de sus secuencias.

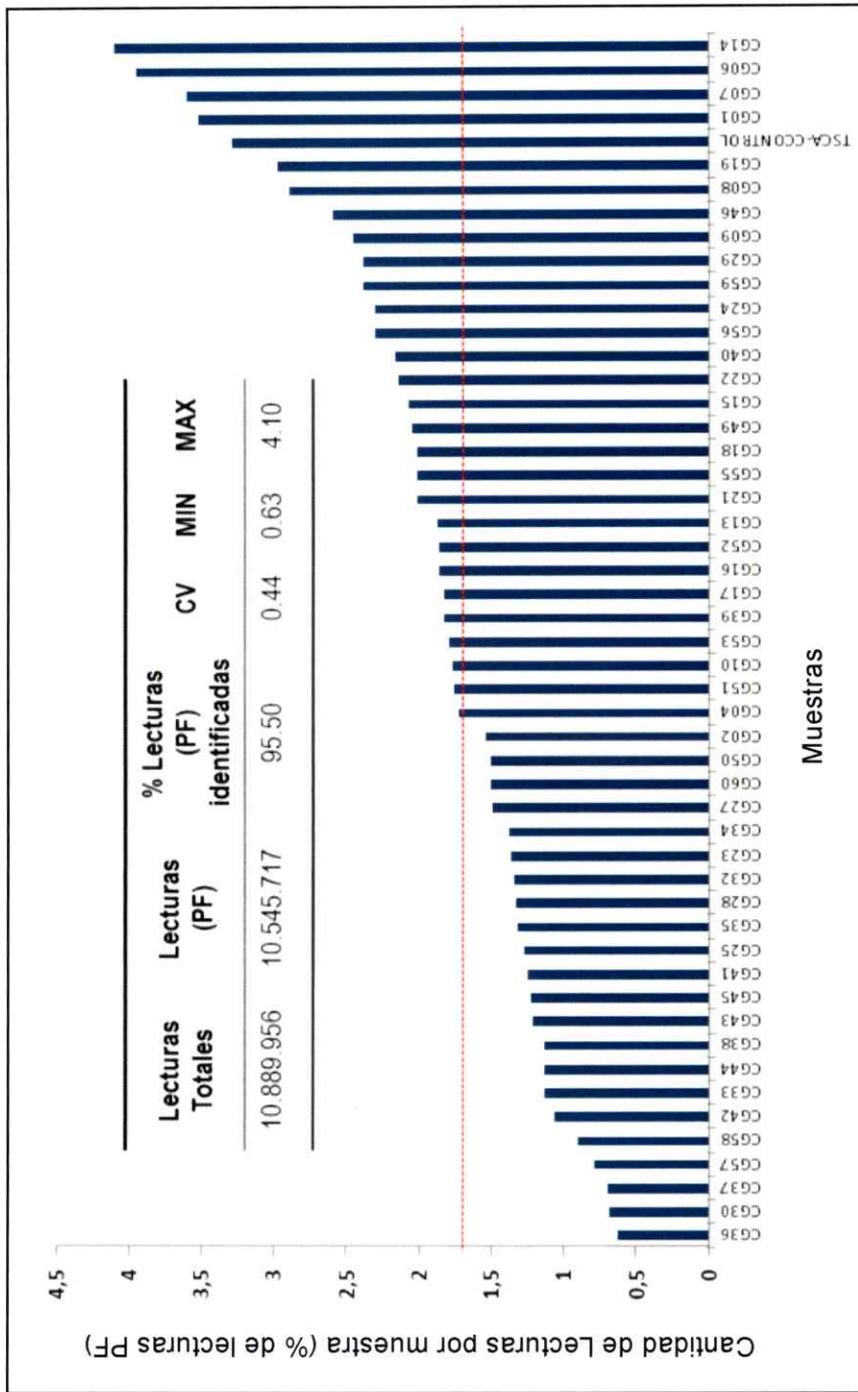


Figura 14. Porcentaje de lecturas identificadas que pasaron el filtro de castidad de Illumina para cada una de las muestras de cáncer gástrico secuenciadas. En el gráfico vemos ordenado de manera creciente el porcentaje de lecturas generadas en la corrida de secuenciación, donde tras el análisis en tiempo real se eliminaron aquellas lecturas que no cumplieren los requisitos de calidad general (Lecturas PF: Lecturas que superaron el filtro). El filtro de castidad de Illumina evalúa la calidad de la imagen en el llamado de bases asignando un valor. Si existe más de una llamada de bases con un valor de castidad de menos de 0,6 en los primeros 25 ciclos, las lecturas no pasan el filtro de calidad. La línea roja indica la media de las lecturas identificadas que superaron los filtros (1.7%). Aproximadamente el 60% de las muestras está por sobre este valor. CV: Coeficiente de variación; MIN y MAX: Valor mínimo y máximo del % de lecturas identificadas (PF)

En relación a las lecturas identificadas del conjunto de muestras en estudio no se observó una distribución homogénea, observándose un rango de lecturas identificadas entre las muestras que va desde 0,5 a 4% del total de lecturas identificadas (10.545.717) (Figura 14). La muestra CG36 es la que posee un menor porcentaje de lecturas identificadas (0.63% del total), mientras que la muestra con mayor porcentaje de lecturas identificadas corresponde a la muestra CG14 (4.10% del total). En base a estos resultados no se encontró una relación entre la calidad previa del ADN genómico utilizado para generar cada bibliotecas y las lecturas identificadas.

En general se espera que el rango de densidades de clusters se encuentre entre los 900 y 1100 K/mm² (Illumina, 2015). Nuestra corrida de secuenciación mostró una densidad promedio de los grupos equivalente a 563 K/mm², lo que puede deberse a la falta de homogeneidad en la concentración inicial de las bibliotecas, pese a que exista una etapa de normalización en la preparación de estas.

Otro parámetro relevante dentro de las métricas de secuenciación corresponde al valor Q30 asignado durante la identificación de cada base añadida a la secuencia (llamado de bases). Este valor es una predicción de la probabilidad de llamada de bases incorrectas. En la corrida de secuenciación se evidenció que un 83.1% de las bases asignadas superaron el filtro de calidad Q30 (Figura 15). Por lo que sobre el 80% de las bases identificadas durante la secuenciación presentaron una probabilidad de ser identificadas erróneamente menor al 0,001 (Q30).

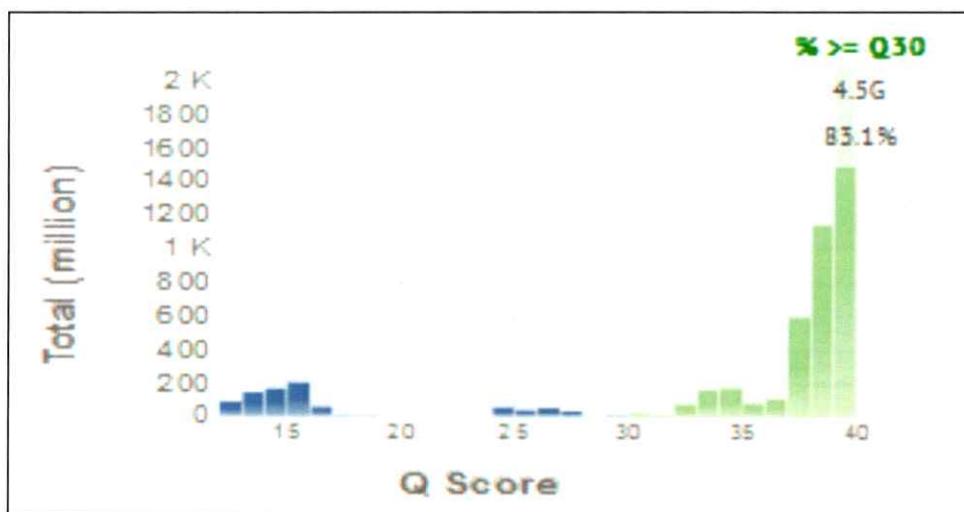


Figura 15. Puntuación de Calidad (Q Score) del llamado de bases para todos los ciclos de secuenciación. La gráfica representa el Q score para el total de bases llamadas durante la corrida de secuenciación. Este valor es una medición de la eficacia del llamado de bases. En esta corrida en particular, un 83.1% de las bases llamadas supero el valor

La corrida de secuenciación incluyó otros genes relacionados a *MICA*, como lo son *MICB* y los genes *ULBP 1-6*, sin embargo los datos generados no fueron analizados en el presente estudio.

Adicionalmente, se determinó la cobertura de los 12 amplicones que cubren el gen *MICA* y se comprobó que la cobertura promedio para las 50 muestras supera los 500x (Figura 16), dando cuenta que en general cada una de las bases de las regiones en estudio se leyeron más de 100 veces y que tal valor asegura que con los datos de secuencia obtenidos es posible realizar análisis informáticos para identificar variantes (Yost y col., 2012).

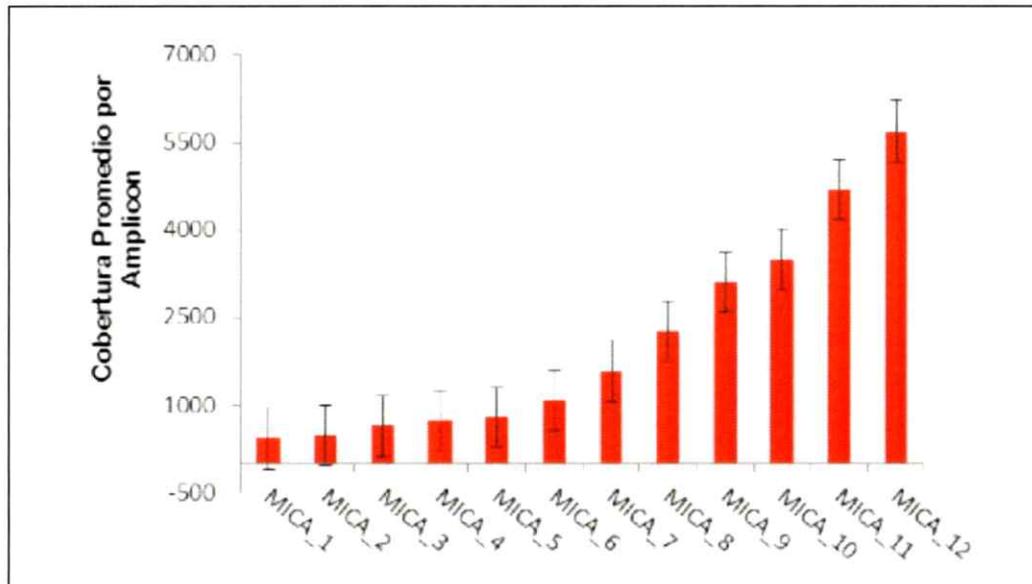


Figura 16. Cobertura promedio por amplicón para el gen *MICA*. La gráfica da cuenta de la distribución de la cobertura promedio de todos los amplicones que cubren el gen *MICA* en estudio, el rango cobertura varía entre los 500 y 5500x.

3.2 Detección de variantes en el gen de *MICA*.

3.2.1 Identificación de variantes genéticas en las muestras en estudio

Mediante el presente estudio de secuenciación masiva del gen *MICA* utilizando la plataforma Illumina MiSeq, se detectaron 62 variaciones de un sólo nucleótido (SNV); de las cuales 23 estaban presentes en sólo una muestra, mientras que 39 estaban en más de una muestra (Figura 17). Se observó además, una distribución homogénea de las variaciones identificadas a lo largo del gen *MICA*, debido a que se observan variaciones en todos los exones en estudio.

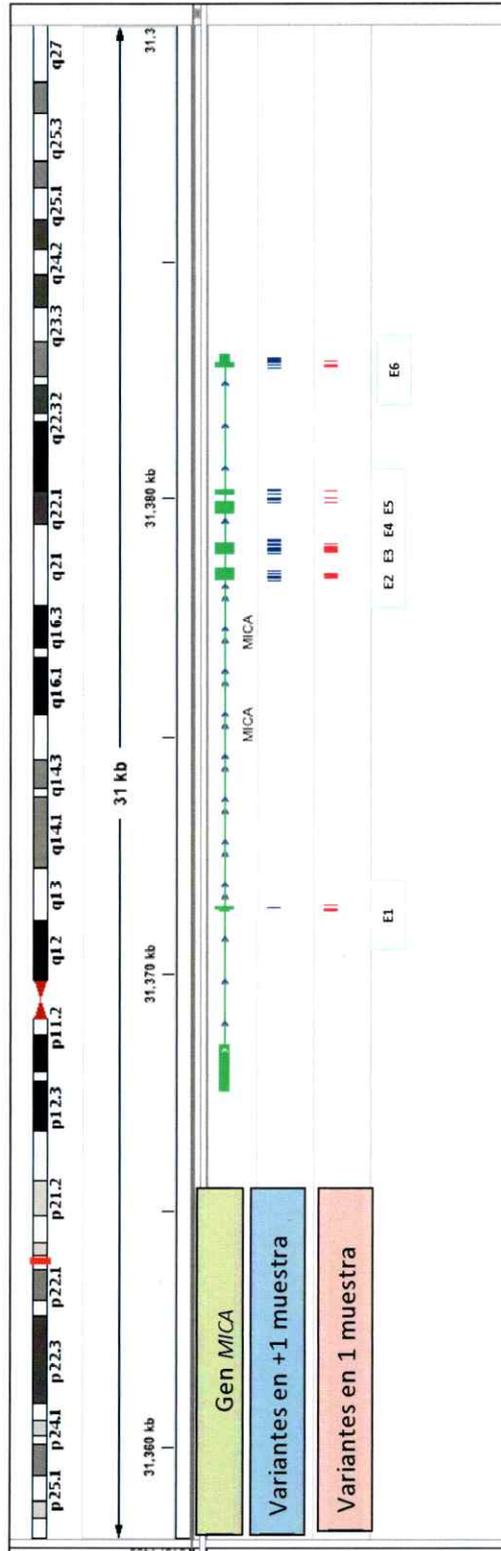


Figura 17. Distribución de las variaciones genéticas para el gen *MICA* detectadas por secuenciación masiva en pacientes con adenocarcinoma gástrico. El gen *MICA* se representa por bandas verdes, las variaciones genéticas detectadas en múltiple muestra (+1) se presentan en azul, mientras que las variaciones presentes en una muestra se presentan en rojo. La distribución de las variantes se presenta de forma uniforme a lo largo de los 6 exones secuenciados del gen *MICA*. Las variaciones fueron visualizadas con el programa IGV versión 2.3.

3.2.2 Caracterización de nuevas variantes del gen *MICA*

De las 39 SNV presentes en más de una muestra del grupo de pacientes analizados, se encontraron tres variaciones en el gen *MICA* que no habían sido descritos previamente en la bases de dato de polimorfismos de un solo nucleótido (dbSNP). Estos tres polimorfismos son:

- **SNV chr6: 31380106 - Identificado en 2 de 50 pacientes.**
- **SNV chr6: 31380114 - Identificado en 7 de 50 pacientes.**
- **SNV chr6: 31380116 - Identificado en 4 de 50 pacientes.**

Cabe señalar que tales variaciones se localizaron en el exón 5 del gen *MICA* (Figura 18).

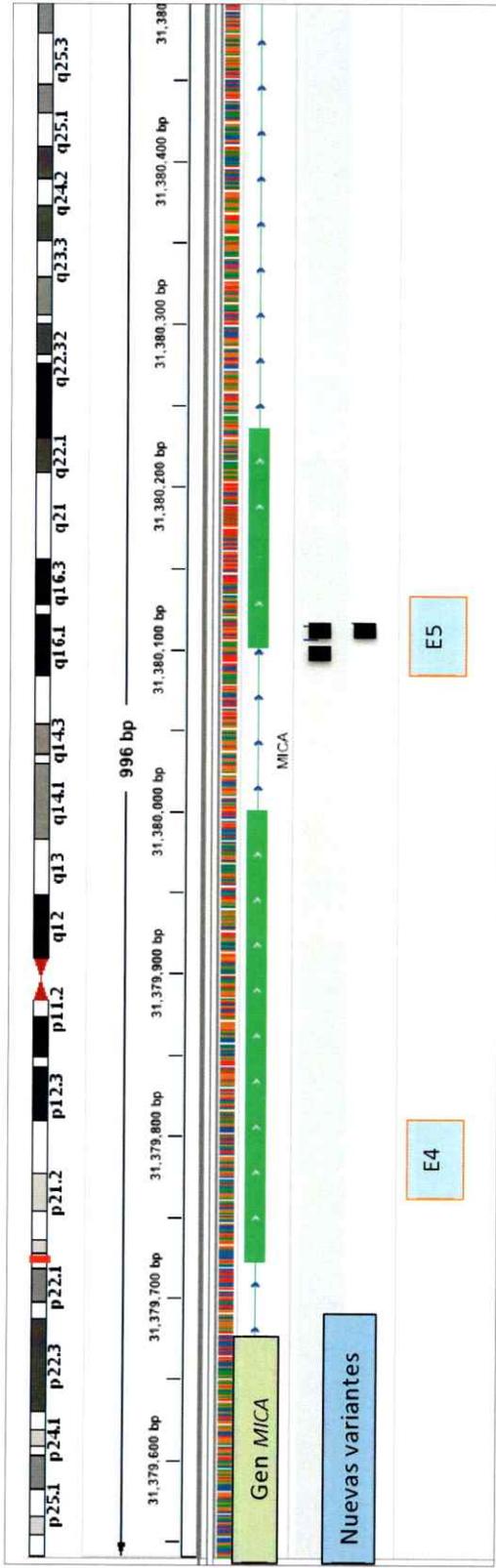


Figura 18. Localización de nuevas variantes detectadas en el gen *MICA*. Del grupo de variaciones detectadas en más de un paciente, tres no se han descrito previamente. Todas las nuevas variaciones se localizan en el exón 5 del gen *MICA*. Las variaciones se visualizaron con el programa IGV versión 2.3.

La variante chr6: 31380106 genera un cambio sinónimo del aminoácido Lisina en la posición 299 de la proteína (K299K). En tanto, la variante chr6: 31380114 genera un cambio no sinónimo en la posición 302 de la proteína, reemplazando Valina por Glicina (V302G) (Figura 19). La variante chr6: 31380116 también genera un cambio no sinónimo en la proteína, cambiando Leucina por Isoleucina (L303I) (Figura 20).

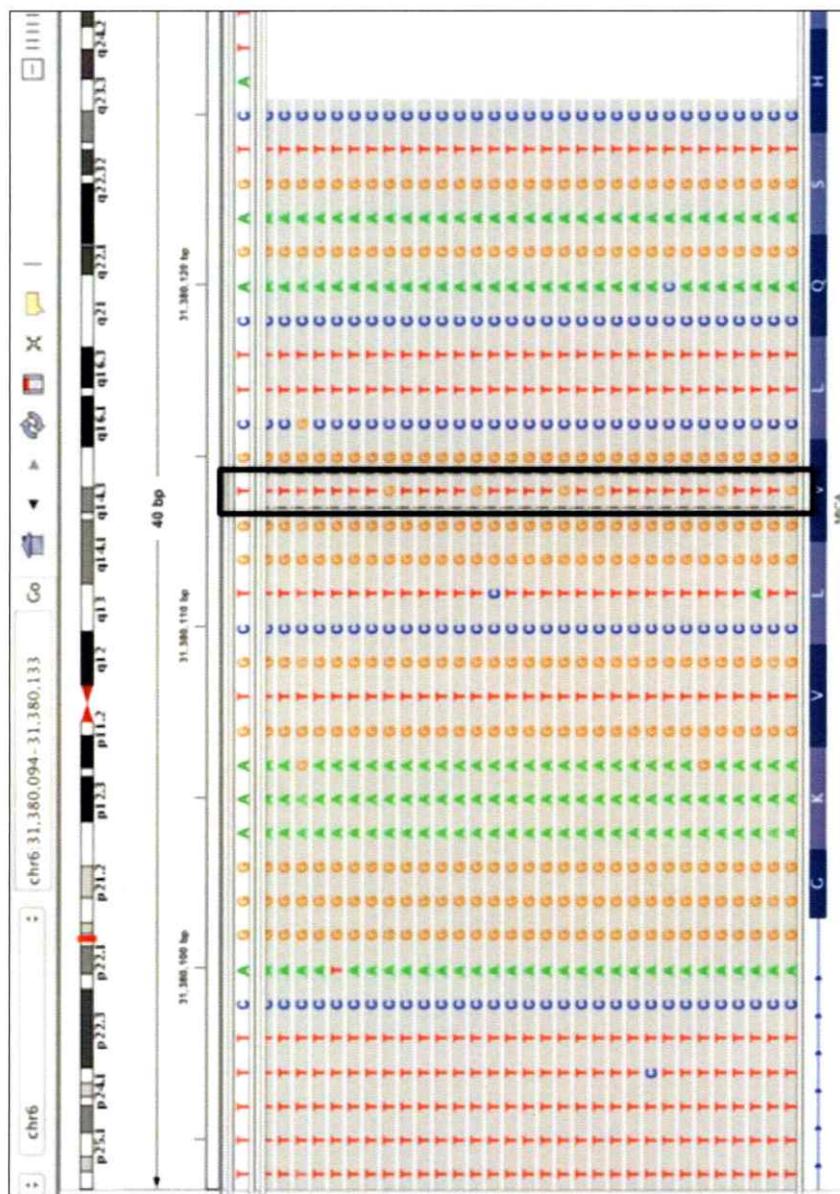


Figura 19. Visualización de nueva variante de MICA (NM_000247:c.944T>G; p.Val302Gly) identificada por secuenciación masiva.
 En el recuadro negro se destaca el cambio de T por G en algunas de las secuencias analizadas. Captura de pantalla de visualización de muestra CG30 en IGV.



En base a los resultados obtenidos por el algoritmo SIFT, sólo la variante chr6: 31380116 tendría un efecto dañino en la proteína MICA, debido a que su SIFT score fue de 0.01, valor inferior al límite 0.05 que predice un cambio de aminoácido como tolerado por la proteína (Tabla 5).

Tabla 5: Efecto de nuevas variantes en la proteína MICA.

Coordinates	Gene Name	Codons	Transcript ID	Protein ID	Substitution	Region	dbSNP ID	SNP Type	Prediction	SIFT Score
380106,1,A/G	MICA	AAA-AAG	ENST00000364810	ENSP00000365394	K299K	EXON CDS	novel	Synonymous	TOLERATED	0.49
380114,1,T/G	MICA	GTG-GgG	ENST00000364810	ENSP00000365394	V302G	EXON CDS	novel	Nonsynonymous	TOLERATED	0.06
380116,1,C/A	MICA	CTT-aTT	ENST00000364810	ENSP00000365394	L303I	EXON CDS	novel	Nonsynonymous	DAMAGING	0.01

3.2.3 Localización y distribución de las variaciones genéticas en el gen *MICA*

De las 62 SNV detectadas, el estudio se centró en variantes presentes en más de una muestra debido a que variantes identificadas en sólo una muestra requieren de estudios adicionales dado que la detección de tales SNVs podría estar dando cuenta de errores propios de la generación de la biblioteca y no de variantes propias del tumor.

Las 39 SNVs recurrentes entre las muestras en estudio se ordenaron en una tabla en que se visualiza la distribución de las variantes de forma creciente en relación a la coordenada cromosómica de las variantes detectadas (Figura 21). En esta tabla con 39 variantes, se identificó la presencia de patrones de variaciones comunes entre diversas muestras.

Se recurrió a la información de una base de datos que contiene marcadores genéticos de tipo SNP presentes en la población Chilena, para determinar si alguna de las mutaciones encontradas era común en la población del país, lo que estaría dando cuenta de variaciones de tipo germinal (heredadas) en los individuos en estudio.

La búsqueda de las variantes identificadas en nuestro estudio dio como resultado que 11 de las 39 variantes se detectaron en la base de datos de ChileGenómico (variantes destacadas en rojo en la Figura 21).

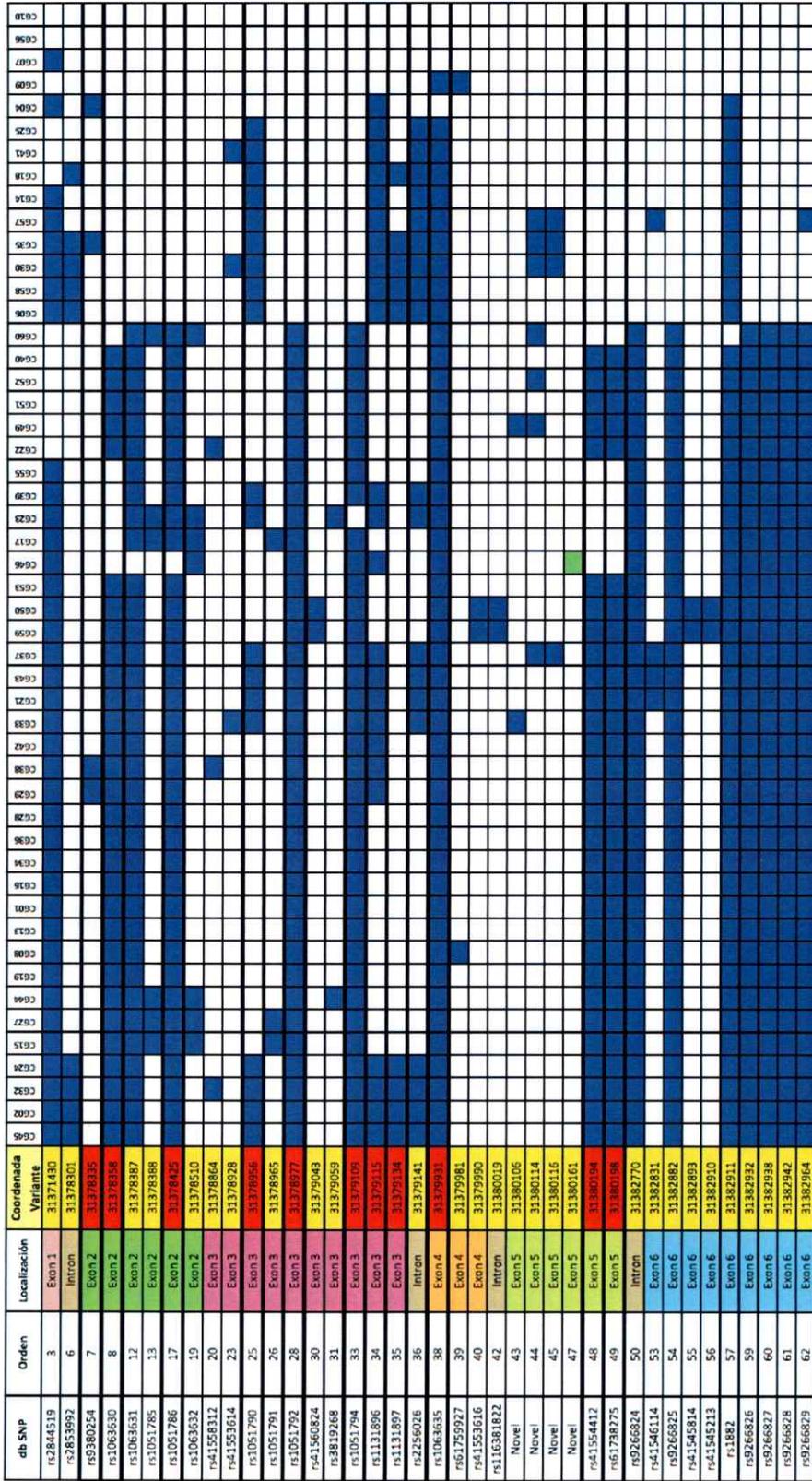


Figura 21. Distribución de los SNV detectados para el gen MICA, e identificación de variantes presentes en una base de datos chilena. En el esquema se ordenan en orden creciente la coordenada cromosómica (amarillo) de las 39 variantes encontradas en más de una muestra, en tanto de izquierda a derecha se ordenan las 50 muestras en estudio. Los cuadros azules representan variantes presentes en más de una muestra y en el recuadro verde se destacan 3 variantes que presentan la misma localización cromosómica. Se observan patrones de bandas en grupos de variantes comunes entre diversas muestras. Se destacan en rojo variantes detectadas en el estudio y que están presentes en la base de datos de ChileGenómico.

3.3 Estudio de mutaciones somáticas en el gen MICA y determinación de su efecto en la proteína.

3.3.1 Identificación de las mutaciones somáticas en el gen MICA.

Las mutaciones somáticas (o adquiridas) son cambios genéticos que ocurren después de la concepción y pueden presentarse en cualquier momento de la vida de una persona. En particular, la cantidad de células del cuerpo que son portadoras de esos cambios depende de cuándo ocurrieron esos cambios, de modo que presentarán una frecuencia alélica diferente de la herencia mendeliana 100 o 50; y se ha sugerido que una frecuencia alélica menor al 35 indicaría mutaciones surgidas durante el desarrollo tumoral (Harismendy y col., 2011). Durante el estudio de mutaciones somáticas en pacientes con adenocarcinoma gástrico, mediante secuenciación masiva, se detectaron 6 variaciones de este tipo en base a sus frecuencias alélicas menores a 35 (Tabla 6); que incluyen a las 3 nuevas variantes previamente analizadas.

Tabla 6: Frecuencias alélicas promedio de las mutaciones somáticas detectadas.

Localización	Variante	Muestra	dbSNP	Prom Freq Alel
31379981	A>A/G	CG08	rs61759927	15.41
		CG09		
31379990	C>C/G	CG50	rs41553616	34.28
		CG59		
31380019	C>C/G	CG50	rs116381822	32.02
		CG59		
31380106	A>A/G	CG33		5.64
		CG49		
31380114	T>T/G	CG30		5.5
		CG36		
		CG37		
		CG49		
		CG52		
		CG57		
31380116	C>C/A	CG30		3.75
		CG36		
		CG37		
		CG57		

De las variaciones somáticas detectadas, se identificó una variante sinónima para la proteína MICA, una variante localizada en la región intrónica entre los exones 4 y 5 del gen en estudio, además de 4 variaciones no sinónimas en la secuencia de la proteína MICA (Tabla 7).

)

Tabla 7 : Mutaciones somáticas para el gen *MICA* detectadas en las muestras en estudio.

Localización	Variante	Num Muestras	Region	ID	Type	AA	AA coord	SIFT	PolyPhen
31379981	A>A/G	2	Exon 4	rs61759927	Missense variant	S/G	291	tolerated(0.11)	benign(0.057)
31379990	C>C/G	2	Exon 4	rs41553616	Missense variant	P/A	294	tolerated(0.24)	possibly damaging(0.556)
31380019	C>C/G	2	intron	rs116381822	Intron variant	-	-	-	-
31380106	A>A/G	2	Exon 5	Novel	Synonymous variant	K	299	-	-
31380114	T>T/G	7	Exon 5	Novel	Missense variant	V/G	302	deleterious(0)	benign(0.394)
31380116	C>C/A	4	Exon 5	Novel	Missense variant	L/I	303	tolerated(0.15)	benign(0.394)

3.3.2 Localización de las mutaciones somáticas

La distribución de las variaciones somáticas en el gen *MICA* se presentó entre los exones 4 y 5, incluyendo una variante localizada en el intrón entre ambos exones (Figura 22).

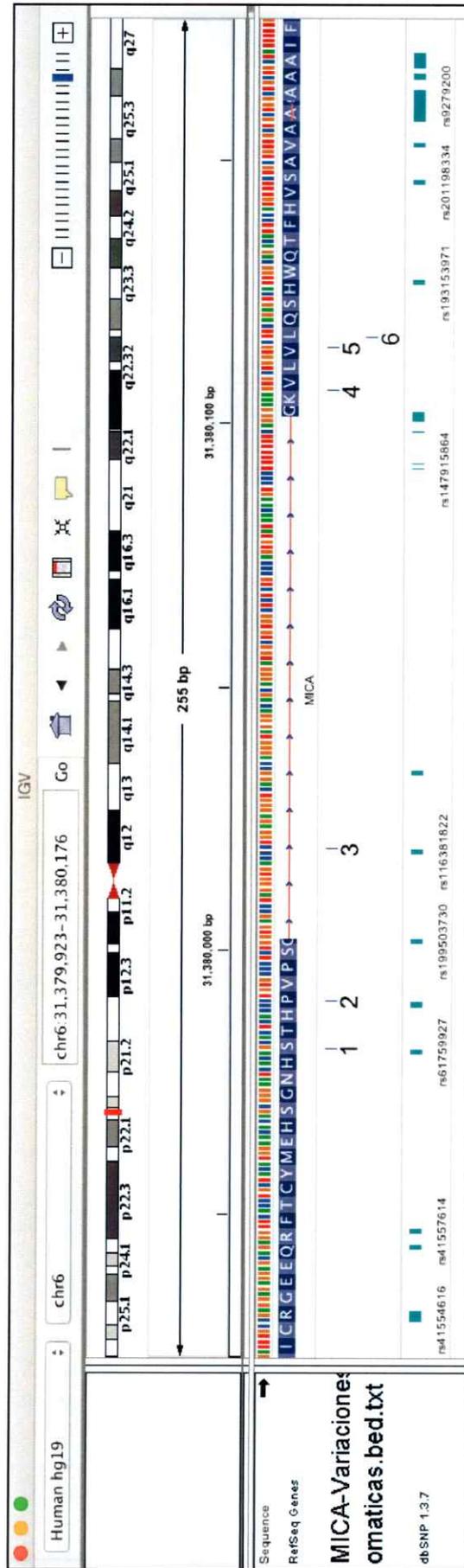


Figura 22: Localización de mutaciones somáticas detectadas para el gen *MICA*. Se detectaron 2 variantes somáticas en el exón 4, 1 variante en el intrón ubicado entre los exones 4 y 5, y 3 variantes somáticas en el exón 5. Las variaciones fueron visualizadas con el programa IGV versión 2.3.

3.3.3 Predicción del efecto de las mutaciones somáticas en la proteína MICA

Debido a que 3 de las 6 variantes somáticas no habían sido descritas previamente en la base de datos dbSNP, estas variaciones no contienen información en la base de datos de SNPeffect. Además, de las 3 restantes una variación se describió como sinónima por lo que no genera cambios en la secuencia de la proteína. Adicionalmente, otra variante se localizó en una región intrónica, por lo que para el presente análisis tampoco se consideró. De este modo, se utilizó solo la variante no sinónima descrita como P294A para la predicción de su efecto en la proteína MICA.

En base a los datos entregados por la base de datos SNPeffect, la variante P294A produce un cambio en la proteína MICA que genera una reducción en la estabilidad de la misma (Figura 23).

Compact overview of VAR_043653

UniProt ID	Gene	Mutation	Mutation type	Disease	OMIM	dbSNP
MICA_HUMAN	MICA	P294A	Polymorphism	-	No OMIM entry	rs41553616

Structure stability summary of VAR_043653

Variant	Stability change	Effect	Structure
P294A	1.3 kcal/mol	Reduced stability	

Figura 23: Efecto de la mutación P294 en la estructura de la proteína MICA. En base a la predicción realizada por la base de datos SNPeffect 4.0, esta variante genera una reducción de la estabilidad de la proteína.

Con el fin de visualizar la mutación P294A en la proteína MICA, se utilizó el programa PyMOL para presentar a partir del archivo PDB (Protein Data Bank) la conformación tridimensional de la proteína MICA y localizar la variación analizada. En particular, la mutación P294A se ubica en el dominio alfa 3 ($\alpha 3$) de la proteína MICA, el cual actúa como nexo entre los dominios de unión al receptor NKG2D y la región transmembrana de la proteína (Figura 24).

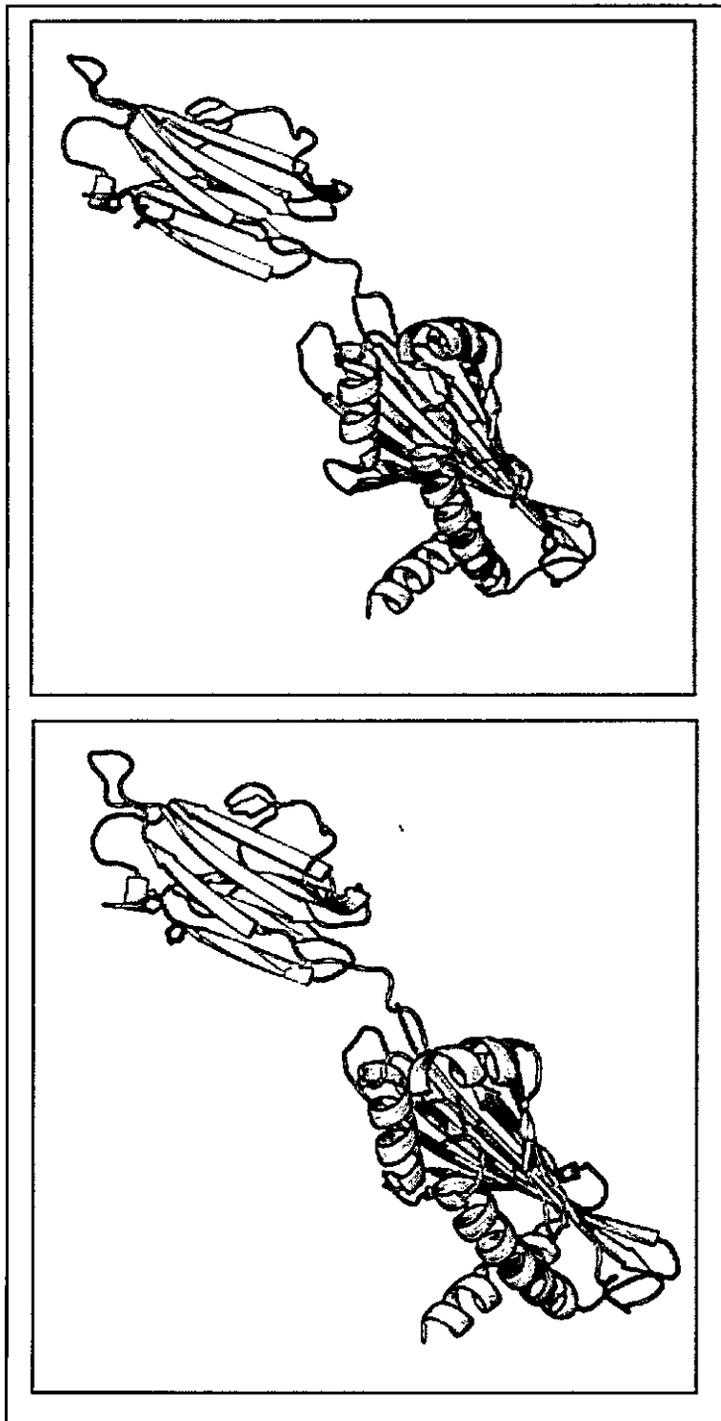


Figura 24. Visualización de la mutación P294A en el dominio $\alpha 3$ de la proteína MICA. Pantalla del programa PyMOL después de cargar el archivo de datos estructurales 1B3J.pdb, correspondiente a los dominios $\alpha 1$, $\alpha 2$ y $\alpha 3$ de la proteína MICA.

4. DISCUSIÓN

En la última década los avances en la investigación genética del cáncer han permitido caracterizar mutaciones propias del tejido tumoral, a menudo referidas como mutaciones somáticas. Una vez que una célula sufre una mutación somática, todas las células que derivan de ella por divisiones mitóticas la heredarán y de este modo, el individuo poseerá un grupo de células con un genotipo diferente al resto. Asimismo, es ampliamente aceptado que los tumores se desarrollan a través de la acumulación de mutaciones somáticas en genes específicos (Zhang y col., 2013).

Dado el hecho que las células derivadas del tejido tumoral expresan de forma simultánea los diferentes ligandos de NKG2D y habiéndose demostrado que el receptor NKG2D participa en la inmunovigilancia y eliminación de células tumorales (Bauer y col., 1999), se esperaría que estas células fueran más susceptibles a la citotoxicidad mediada por linfocitos NK, NKT, T CD8+ o T $\gamma\delta$, resultando en el control del crecimiento tumoral y un pronóstico más favorable para la enfermedad. Sin embargo, estos tumores han podido progresar, lo cual implicaría que las interacciones entre el receptor NKG2D y sus ligandos estarían alteradas en el microambiente tumoral. Dado que, si bien los ligandos de NKG2D están presentes en las células derivadas del tumor, podrían presentar alteraciones en su secuencia que generasen cambios conformacionales en el ligando o que promovieran la liberación de ligandos solubles.

Debido a lo anterior, en el presente estudio se evaluaron variaciones genéticas de tipo SNV en el gen *MICA* que actúa como ligando de NKG2D, donde tales variaciones

estarían dando cuenta de mutaciones somáticas en pacientes con CG y se analizaron los efectos de estos cambios en la función de la proteína.

4.1 Detección de variaciones en el gen *MICA*

MICA es un gen altamente polimórfico, a la fecha se conocen 91 variantes alélicas (datos de base de datos IMGT) y si bien no se conoce el papel que la mayoría de estos polimorfismos ejercen sobre la función de la molécula, en algunos casos se ha descrito que estas variaciones afectan a la expresión de *MICA* o a su capacidad de activación de las células citotóxicas del sistema inmune (González y col., 2006)

Con el fin de detectar mutaciones somáticas en regiones codificantes del gen *MICA*, se secuenció un grupo de 50 muestras tumorales provenientes de pacientes con adenocarcinoma gástrico.

Las 50 muestras fueron agrupadas en 5 grupos de acuerdo a criterios de integridad de su material genético (Figura 8), esperando que tal agrupación se asociara de manera directa en que muestras que presentaran fragmentos mayores a 7.000 pb tuviesen una mejora en la generación de bibliotecas, no obstante, los resultados de secuenciación sugieren que la calidad de la biblioteca habría sido independiente del grupo de calidad del ADN genómico asignado a la muestra.

En este contexto es importante destacar que en relación a las lecturas identificadas y la clasificación de las muestras no se observó una distribución homogénea a través de los grupos, ya que en particular dos muestras que se encontraban en el mismo conjunto de acuerdo a la integridad de su ADN genómico (Grupo II) presentaron los valores extremos de la distribución de valores de lecturas identificadas luego de la

secuenciación, los que variaron entre el 0,5 a 4% (Figura 14). El menor porcentaje de lecturas identificadas lo presentó la muestra CG36, mientras el mayor porcentaje de lecturas identificadas correspondió a la muestra CG14, por lo tanto, en base a estos resultados no se encontró una relación directa entre la integridad del ADN genómico utilizado para generar cada biblioteca y las lecturas identificadas. Ante esto, es probable que las variaciones de porcentaje de lecturas identificadas entre muestras se deban al procedimiento de generación de la biblioteca para cada muestra, en particular a la etapa de purificación de fragmentos de ADN utilizando beads magnéticos; donde una manipulación no óptima conllevaría a la pérdida de ADN y reducción del material genético disponible para la generación de los grupos para cada muestra (Bronner y col., 2013)

Por otra parte, los resultados obtenidos de la secuenciación indicaron que el 95.5% de las lecturas generadas pasaron los filtros de castidad, indicadores de la buena calidad de los grupos generados durante la etapa de amplificación (Figura 14). De esta forma se eliminaron aquellos datos menos fiables, los que podrían conllevar la detección de falsos positivos en los análisis de datos posteriores (Minoche y col., 2011).

Otro parámetro relevante dentro de las métricas de secuenciación corresponde a la puntuación de calidad (Q30) del llamado de bases para todos los ciclos de secuenciación. Este valor es una medición de la calidad del llamado de bases y en particular corresponde a la probabilidad de una llamada de bases incorrecta por cada 1000 bases agregadas (Trivedi y col., 2014; Rutvisuttinunt y col., 2013). A partir de los resultados de secuenciación, se evidenció que un 83.1% de las bases asignadas

superaron el filtro de calidad Q30 (Figura 15); lo que estaría dando cuenta de una óptima calidad de las lecturas generadas debido a que en relación a las características del cartucho de la plataforma de secuenciación utilizada, se esperaría un puntaje de Q30 superior al 75%. (Rutvisuttinunt y col., 2013). En base a ello, es posible asegurar una buena calidad de los datos de secuencia generados durante la secuenciación y que estos son aptos para análisis informáticos posteriores.

Un punto importante a destacar es que el proyecto inicial consistía en secuenciar los exones 2,3,4 y 5 del gen *MICA*, junto a los otros 7 ligandos de NKG2D (*MICB* y *ULBP 1* al 6). Sin embargo, luego de analizar el archivo Manifiesto y mapear los amplicones para la corrida de secuenciación, se hizo evidente la presencia de amplicones adicionales que estaban mapeando exones anexos de los genes en estudio, además de pseudogenes como *MICC* y *MICD*. Esto se podría explicar debido a la similitud de secuencia que presentan los genes de la familia MIC (Chen & Gyllensten, 2014), lo que implicaría que las sondas dirigidas a regiones de interés de los genes *MICA* y *MICB* hibridaran de manera inespecífica a regiones de estos pseudogenes. Adicionalmente, la detección de nuevos amplicones que estén cubriendo regiones incluidas en el estudio, podría deberse a problemas detectados por Illumina durante la generación misma de las sondas y que para reducir errores de lectura posteriores, se incluyeran amplicones adicionales para sobre leer tales regiones. Por lo anterior, se podría suponer que la inclusión de nuevos amplicones conllevaría a la disminución de reactivos de secuenciación disponibles para las sondas que mapean regiones del estudio, debido a la inclusión de nuevos amplicones en el proceso. No obstante, es importante destacar que luego de la secuenciación se

evidenció que la mayoría de los amplicones adicionales al estudio presentaron una nula o baja cobertura, por lo cual no afectaron la química de la secuenciación de las regiones estudiadas. Cabe señalar que para el gen *MICA*, se agregaron 4 amplicones: uno que mapeo el exón 1, dos mapearon el exón 3 y uno mapeó el exón 6, por tal razón el presente estudio generó datos de secuencia de los 6 exones que conforman el gen *MICA*, lo que permitió detectar variantes en regiones que no fueron seleccionadas en el estudio inicial.

El análisis bioinformático de los datos de secuenciación se realizó siguiendo el flujo de trabajo Amplicon TruSeq (Illumina) (Figura 5) y para el llamado de variantes se utilizó el algoritmo Somatic Variant Caller, recomendado por Illumina para el análisis de datos generados con el kit utilizado en este proyecto (TruSeq Custom Amplicon, Illumina Inc.). En el archivo de formato de llamadas de variante (VCF) se detectaron 62 variaciones de un sólo nucleótido distribuidas de forma homogénea en los 6 exones del gen *MICA*; 23 variantes estaban presentes en sólo una muestra, mientras que 39 estaban en más de una muestra.

En el estudio se detectaron tres nuevas variantes y los resultados entregados por el algoritmo SIFT sugieren que la variante chr6: 31380116 tendría un efecto dañino en la proteína MICA (Tabla 5). Este algoritmo se basa en el grado de conservación de los aminoácidos en alineamiento de secuencia derivados de secuencias estrechamente relacionadas y recogidas a través de PSI-BLAST (Kumar y col., 2009), por tal razón, la variante detectada podría tener un efecto negativo significativo en la función de la proteína MICA.

Cabe destacar que la muestra CG10 no presentó variantes de tipo SNV (variante de un sólo nucleótido) para el gen *MICA*, lo que se podría explicar por errores tanto en el proceso de preparación de la biblioteca, así como a la integridad del ADN genómico utilizado en el proceso. Sin embargo, al analizar los datos del gen *MICB* secuenciado en paralelo a *MICA* para la misma muestra, se observa que sí presenta variantes de tipo SNV, de modo que este antecedente nos podría sugerir que la ausencia de variantes en *MICA* podría deberse a que si bien se detectaron SNV, estos fueron filtrados debido a su baja calidad durante el flujo de análisis bioinformático, por lo que no se detectaron al generar el archivo VCF. Estudios recientes del laboratorio han detectado que los filtros utilizados por el flujo de trabajo desarrollado por Illumina son estrictos en relación a la calidad de secuencia para la identificación de variantes, y que un re-análisis de los mismos datos con flujos de trabajos con GATK y MUTEC han permitido detectar variantes que previamente no habían sido detectadas por el flujo de trabajo de Illumina; por tal razón se hace necesario revisar nuevamente los datos brutos de secuencias para el gen *MICA* de la muestra CG10 con otros algoritmos de modo de poder detectar variantes en esta muestra.

Del mismo modo es importante destacar que para la muestra CG46 se detectaron 3 variantes con la misma coordenada cromosómica (cuadro verde en Figura 21). Tal fenómeno podría explicarse por la presencia de una variación genética de tipo Inserción o Delección (Indel) cercano a esa posición; tal mutación que involucre a más de un nucleótido, podría generar secuencias con corrimientos de marco de lectura en relación al genoma de referencia, las que no son detectadas por el algoritmo SVC debido a que

sólo detecta variaciones de una base de longitud. Y esta sería la razón por la que el flujo bioinformático detectó más de una variante para una única posición en el gen.

4.2 Mutaciones somáticas en el gen *MICA* y su efecto en la proteína

En base a la Figura 21 se evidenció la presencia de patrones de SNV a lo largo del gen *MICA* que eran comunes entre varias muestras, en particular se observaron bloques de variaciones que se formaban por agrupaciones de variantes adyacentes y que se presentaban como líneas compactas a través de la figura. Estos bloques de variantes podrían estar dando cuenta de haplotipos, es decir, conjuntos de polimorfismos de un sólo nucleótido (SNPs) presentes en un cromosoma particular que están estadísticamente asociados y que se heredan en bloque; y que se relacionan con variaciones presentes en la población heredándose de padres a hijos (Goldstein & Cavalleri, 2005).

Basado en lo anterior, se procedió a descartar los SNPs relacionados a posibles haplotipos de modo de eliminar variantes de tipo germinal (heredadas) en las muestras en estudio. Para ello se utilizaron datos presentes en la base de datos ChileGenómico, el cual corresponde a un muestreo poblacional de SNP (proyecto financiado por FONDEF D10I1007: “Genómica de la población chilena: Obtención de perfiles genéticos necesarios en investigación clínica, salud pública y medicina forense”). Al buscar las 39 variantes de tipo SNV detectadas en el estudio en la base de datos chilena, se identificó que 11 de las variaciones que se encontraban en bloques correspondían a SNP poblacionales chilenos. De este modo al descartarlos (cuadros rojos en la Figura 21) se seleccionaron 28 SNV candidatos para ser considerados variaciones propias del tumor (somáticas).

A continuación, se analizaron los datos de secuenciación en bases a la frecuencia alélica de los 28 posibles SNV somáticos. En particular se tiene que las variaciones heredadas (germinales) presentan una frecuencia alélica de 50% cuando se encuentran en condición de heterocigosis y de 100% cuando son homocigotas, sin embargo, durante el desarrollo tumoral ocurren variaciones somáticas las que presentarán una frecuencia alélica menor al 50 % ya que debido a las sucesivas divisiones mitóticas de la célula que la carga sólo una porción del tejido la presentará (Harismendy y col., 2011).

Al analizar la frecuencia alélica se identificaron 6 variantes somáticas distribuidas en los exones 4, 5 y en el intrón entre ambos. Con el fin de determinar el efecto de estos SNV en la proteína MICA es importante destacar que si bien la variante rs 116381822 presente en el intrón no tendrían un efecto directo en la conformación de la proteína, podría estar implicada en el proceso de splicing (Wang y col., 2008); lo que podría sugerirla como objeto de nuevos estudios. Adicionalmente, una de la variantes detectadas genera un efecto sinónimo en la posición 299 en la proteína por lo que tampoco se consideró en los análisis posteriores.

Si bien 4 de las 6 variantes somáticas tendrían un efecto en la proteína MICA (Tabla 7), dos de ellas (V302G y L303I) no han sido descritas previamente en base de datos de SNP (dbSNP) por lo que no existe información en bases de datos referente a su efecto en la proteína.

Finalmente, se procedió a analizar las variantes presentes en bases de datos, para ello se utilizó la base de datos SnpEff 4.0 que contiene información de SNPs codificantes no sinónimos y analiza el efecto fenotípico de variaciones en genes humanos. La versión actual (4.0), incluye datos de estabilidad de proteínas, integridad de sitios funcionales,

agregación de proteínas, entre otros (De Baets y col., 2012). En particular, la variante somática P294A presentó un efecto en la proteína MICA de acuerdo a SNPEffect (Figura 23). Los datos sugieren que el cambio de una prolina por una alanina en la posición 294 generó una reducción de estabilidad de la proteína, donde el cambio es cualitativamente importante debido a que los requerimientos estructurales de los enlaces peptídicos generados por prolina están directamente relacionados con la estructura terciaria de la proteína; en particular la ausencia de anillo por parte de la alanina podría generar una diferencia conformacional que implicaría cambios en los parámetros cinéticos y de estabilidad de la proteína (Street y col., 2005).

El cambio conformacional generado por la variación P294A puede verse en una representación tridimensional del ligando MICA con el programa PyMol (Figura 24). Es importante destacar que esta mutación somática se ubicó al final del exón 4, el cual codifica para el dominio extracelular alfa 3 de la proteína MICA y que se encuentra adyacente al dominio transmembrana. En este punto es importante destacar que la posición de la variación se localiza en el último residuo del dominio alfa 3, el que está unido a un linker de 6 aminoácidos que a su vez lo une al dominio transmembrana de la proteína. Tal localización de la variante podría ser un lugar de posible corte proteolítico de MICA por parte de metaloproteinasas (Chen & Gyllensten, 2014), donde el cambio conformacional generado por el cambio no sinónimo de prolina por alanina dejaría disponible un sitio de corte por estas enzimas. De este modo la generación del ligando MICA soluble podría estar dando cuenta de mecanismos de evasión tumoral por parte de las células transformadas, donde ligandos que no están unidos a la célula se estarían uniendo a NKG2D en la superficie de células NK, y de este modo actuarían como

inhibidor competitivo bloqueando el reconocimiento de moléculas MICA unidas a la membrana, y enmascarando la posible unión de un ligando efectivamente anclado a la célula tumoral (Coudert y col., 2006).

Las características clínico patológicas de los pacientes que presentaron la variante somática con efecto en la proteína MICA son las siguientes: el paciente CG50, que corresponde a un varón de 54 años con un tumor localizado en la curvatura menor del estómago y un TNM de 3; en tanto, el paciente CG59 es un varón de 69 años con un tumor localizado en la zona antral del estómago y con un TNM de 4. Estos datos sugieren que la detección de variantes somáticas en su ADN se podría relacionar con el avanzado estado de la enfermedad, debido a que se ha reportado que procesos celulares de control internos están muy alterados debido a aumentos en la carga de mutaciones somática con la edad, con consecuencias para el envejecimiento, el cáncer y la neurodegeneración (Kennedy y col., 2012).

En base a ello, se ha reportado que la progresión tumoral requiere de un transcurso de muchos años y se ha demostrado que desde el punto de vista molecular, el cáncer es un proceso de múltiples etapas que requiere varios eventos de mutación somática y de selección clonal, que producen a lo largo del tiempo variantes en la descendencia celular con propiedades de crecimiento cada vez más agresivas (Kennedy y col., 2012). Además, el análisis genético ha demostrado que el cáncer es un proceso secuencial que implica la acumulación de sucesivas mutaciones en uno o varios genes de distintos tipos (oncogenes, genes supresores, genes de susceptibilidad, etc) y la subsiguiente selección clonal de las células portadoras de las mismas, que conduce a que estas células alteren su

comportamiento incrementando su capacidad proliferativa, de invasión y llegando eventualmente a causar la muerte del organismo del que proceden (Campisi, 2013).

En resumen, los resultados obtenidos en este estudio nos permiten corroborar la hipótesis propuesta. Se observó la presencia de variaciones somáticas de tipo SNV en la secuencia del gen *MICA*, cuyo efecto podría estar relacionado con la liberación de ligandos solubles y progresión tumoral. Sin embargo, se requieren análisis más exhaustivos para poder determinar el real efecto de las variantes en la conformación de la proteína; además de estudiar el efecto de las variantes somáticas que no han sido descritas previamente (nuevas) y determinar su efecto sobre el ligando MICA en cáncer gástrico.

Es importante destacar que si bien *a priori* se podría pensar que los SNV con mayor efecto en la proteína se ubicarían en la zona de interacción con el receptor; nuestros resultados sugieren que cambios conformacionales en dominios distantes al sitio de interacción podrían dar cuenta de zonas de posibles cortes proteolíticos y la consecuente generación de ligandos solubles (Obuchi y col., 2001). Adicionalmente, para poder afirmar con certeza la identificación de variaciones somáticas es necesario secuenciar la sangre de los pacientes en estudio, de modo de poder descartar las variaciones germinales para cada individuo. Este procedimiento podría realizarse mediante secuenciación de tipo Sanger, específica para las coordenadas cromosómicas de las variaciones somáticas detectadas.

La identificación de mutaciones somáticas desde datos de secuenciación masiva es un reto debido a la dificultad de distinguir los verdaderos acontecimientos somáticos de los artefactos derivados de la PCR, errores de secuencia o de la mala asignación, por tal

razón para tener una mayor confianza de los resultados obtenidos se hace necesario analizar la celularidad tumoral del tejido en estudio, es decir, incluir en nuestro análisis el factor de pureza tumoral y sub-clonalidad de las muestras analizadas, y de este modo identificar de forma certera verdaderos acontecimientos somáticos de aquellas variantes presentes en línea germinal (Kassahn y col., 2013).

Por otra parte, la detección de posibles haplotipos en el gen *MICA* se podrían correlacionar con que este gen esta localizado en el brazo corto del cromosoma 6, perteneciendo al complejo principal de histocompatibilidad humano (HLA) (Zwirner y col., 1998), donde una de las características fundamentales de este grupo de genes es su alto grado de polimorfismo y la presencia de desequilibrio de ligamiento. Así, en poblaciones, ciertas combinaciones de alelos se heredan juntos más frecuentemente de lo esperado debido al azar y conjuntos de polimorfismos heredados podrían conducir a diferentes haplotipos. Esto contribuiría a diferencias en la respuesta inmunitaria entre individuos (Tian y col., 2001); jugando un papel importante en los trasplantes de órganos y la susceptibilidad a ciertas enfermedades, entre los que destaca el cáncer. En este punto, es importante señalar que los resultados obtenidos podrían servir como antecedentes para desarrollar un nuevo proyecto, en el ámbito de la genética poblacional, que investigue la función de los haplotipos presentes en el gen *MICA* como posibles marcadores moleculares o predictores de riesgo a padecer cáncer gástrico.

Finalmente, los resultados obtenidos en la presente Memoria nos permiten postular que la presencia de mutaciones somáticas de tipo SNV en la secuencia del gen *MICA* generan variaciones no sinónimas en la proteína con un posible efecto en su conformación, promoviendo posiblemente la liberación soluble del ligando MICA y que

este podría ser uno de los mecanismos por el cual el adenocarcinoma gástrico escapa de la vigilancia inmune.

Adicionalmente, la utilización de secuenciación masiva supone un gran desafío conjunto entre investigadores y médicos, de modo de poder implementar protocolos de prevención y detección temprana de CG.

5. CONCLUSIONES

Los resultados de nuestro estudio demuestran que es posible detectar variaciones genéticas de tipo SNV en el gen *MICA* a partir de secuenciación masiva de ADN genómico proveniente de pacientes con adenocarcinoma gástrico.

Se observó la presencia de variaciones presentes en una única muestra, así como de SNV comunes entre muestras; sin embargo, para nuestro estudio de detección de mutaciones somáticas, sólo analizamos los SNV presentes en más de una muestra, ya que variantes identificadas en sólo una muestra requieren de estudios adicionales dado que la detección de tales SNVs podría estar dando cuenta de errores propios de la generación de la biblioteca y no de variantes propias del tumor.

Tres de las variantes detectadas no habían sido descritas previamente en la base de datos dbSNP. No obstante, sólo una de ellas tendría un efecto significativo en la proteína *MICA* respecto a su conformación normal.

La secuenciación masiva del gen *MICA* permitió identificar diversas variaciones somáticas, las que estarían dando cuenta de eventos de mutación propios del desarrollo tumoral.

Se detectó que algunos tumores en estadios tardíos (III y IV) presentaron mutaciones somáticas que generaron un efecto sobre la proteína *MICA*, lo que podría estar asociado a la progresión del tumor.

Futuros estudios podrían centrarse en los haplotipos detectados en las secuencias del gen *MICA*, de modo de analizar su función como posibles marcadores moleculares o predictores de riesgo a padecer CG.

6. ANEXOS

Anexo 1



FACULTAD DE MEDICINA
UNIVERSIDAD DE CHILE

HOSPITAL CLINICO
UNIVERSIDAD DE CHILE



CONSENTIMIENTO INFORMADO PACIENTES CON DIAGNÓSTICO DE CÁNCER GÁSTRICO

"Estudio de factores genéticos, determinantes de virulencia de *Helicobacter pylori*
y su relación con cáncer gástrico"

INVESTIGADOR PRINCIPAL:

Patricio González Hormazábal
Facultad de Medicina, Universidad de Chile
Fono: 9786845
e-mail: ppgnzalez@med.uchile.cl

Antes de tomar la decisión de participar en la investigación, lea atentamente este formulario.

Mi médico tratante me ha informado que padezco cáncer gástrico. Me someteré a cirugía como parte del tratamiento de mi enfermedad, y se me ha invitado a participar del estudio "Factores genéticos, determinantes de virulencia de *Helicobacter pylori*, y su relación con cáncer gástrico".

El Investigador principal del estudio me informó que está realizando un estudio de la relación entre factores genéticos y la infección con determinados tipos de bacteria *Helicobacter pylori*. Las conclusiones de este estudio serán de utilidad para comprender mejor las causas del cáncer gástrico y la posibilidad de identificar aquellas personas que tengan alto riesgo de desarrollar esta enfermedad con el fin de prevenir su aparición.

Se me ha informado que mi participación en calidad de paciente afectado de cáncer gástrico consiste en (a) responder una encuesta que recopila información de mis datos personales e historia familiar de cáncer, (b) donar 3 biopsias gástricas que se obtendrán desde la pieza quirúrgica que se me extirpará durante la cirugía, desde las cuales se obtendrá material genético (DNA) de la bacteria *Helicobacter pylori* y se analizará el efecto de los factores genéticos analizados, (c) donar 1 biopsia del tumor que se obtendrá desde la pieza quirúrgica que se me extirpará durante la cirugía, desde la cual se analizará el efecto de los factores genéticos analizados, y (d) donar una muestra de sangre tomada desde una vena de su antebrazo, desde la cual se obtendrá material genético (DNA) y suero.

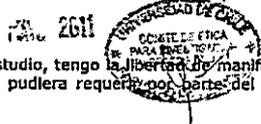
Se me informó que mi participación en este estudio no me representará gastos adicionales, ni a mí ni a mi sistema de salud, a lo ya cancelado por las prestaciones que se me han realizado por orden de mi médico tratante.

Se me ha informado que la toma de muestra de sangre que se me realizará podría producir molestas tales como dolor, hemorragia, o reacción alérgica por el antiséptico utilizado. Sin embargo, la muestra será tomada por una Enfermera Universitaria con experiencia en este tipo de procedimientos, quien tomará todas las precauciones para minimizar las posibilidades que esto ocurra.

Se me ha informado que las biopsias gástricas se obtendrán desde la pieza quirúrgica que se me extirpará durante la cirugía, por lo cual la toma de biopsias está exenta de riesgos.

Se me ha informado que las muestras y toda mi información personal serán identificadas con un código numérico para su uso actual o futuro. Los resultados individuales serán anónimos, se mantendrán en reserva y sólo serán de uso y conocimiento del Investigador.

Se me ha informado que mi participación en este estudio no me otorga ningún beneficio médico o compensación económica alguna. Sin embargo, se me comunicó que la donación voluntaria de muestras e información destinadas a realizar esta investigación, podrá ser beneficiosa para futuras generaciones.



Estoy en conocimiento que, en caso de decidir no participar en este estudio, tengo la libertad de manifestar mi decisión y ésta no cambiará en nada las prestaciones normales que pudiera requerir por parte del Hospital Clínico de la Universidad de Chile.

Se me ha informado que el Investigador Principal, la Universidad de Chile y el Hospital Clínico de la Universidad de Chile velarán por que se proteja la confidencialidad de mi identidad, registro médico y resultados de esta investigación.

Por último he leído este formulario y se me permitió realizar todas las preguntas que consideré pertinentes y de mi interés, las que fueron contestadas a mi entera satisfacción. Se me ha permitido consultar con mi médico de familia o pedir la opinión de otro profesional en relación a mi participación en este estudio. Entiendo que se me dará copia de este documento.

Consiento en participar de la Investigación "Estudio de factores genéticos y su relación con cáncer gástrico".

Acepto que parte de las muestras se congelen y almacenen para estudios futuros que puedan surgir en relación al estudio de factores genéticos y su relación con cáncer gástrico. SI NO

¿Desea conocer los resultados de la Investigación?. SI NO

Nombre del participante: _____

R.U.T.: _____ Firma : _____

Nombre del Investigador: _____

R.U.T.: _____ Firma : _____

Nombre del delegado del Director del Hospital (Ministro de Fé): _____

R.U.T.: _____ Firma : _____

Santiago, ____ de ____ de 20__

Cualquier duda o consulta que tenga sobre este estudio, o experimenta cualquier problema, siéntase libre de comunicarse con el Investigador.

Frente a cualquier duda respecto a aspectos legales y éticos de este estudio, siéntase libre de comunicarse con: Dr. Juan Jorge Silva, Presidente del Comité de Ética, Hospital Clínico de la Universidad de Chile.

Av. Santos Dumont 999, Santiago. Fono 9789008.

Dr. Manuel Oyarzún, Presidente del Comité de Ética de la Investigación en Seres Humanos.

Facultad de Medicina, Universidad de Chile, Av. Independencia 1027, Santiago. Fono 9786923.

Con copia a: - Participante - Investigador

Anexo 2



UNIVERSIDAD DE CHILE
FACULTAD DE MEDICINA
COMITÉ DE ÉTICA DE INVESTIGACIÓN EN SERES HUMANOS



31 MAYO 2011

1/2

ACTA DE APROBACIÓN PROYECTO DE INVESTIGACIÓN SERES HUMANOS

Con fecha 24 de mayo de 2011, el Comité de Ética de Investigación en Seres Humanos de la Facultad de Medicina, Universidad de Chile, integrado por los siguientes miembros:

Dr. Manuel Oyarzún G., Médico Neumólogo, Presidente
Sra. Marianne Gaudlitz H., Enfermera, Mg. Humanidades, Vicepresidente
Dr. Hugo Amigo C., Ph. D., Especialista en Salud Pública
Dr. Leandro Blagini A., Médico Internista
Dra. Lucía Cifuentes O., Médico Genetista
Sra. Nina Horwitz C., Sociólogo, Mg. Bioética
Dra. María Eugenia Pinto C., Médico Infectólogo
Sra. Claudia Marshall F., Representante de la comunidad

Ha revisado el Proyecto de Investigación titulado: "ASSOCIATION OF CYTOKINE POLYMORPHISMS AND OF GENETIC DETERMINANTS OF *H. pylori* VIRULENCE FACTORS WITH GASTRIC CANCER RISK" y cuyo Investigador responsable es el Dr. **Patricio González H.**, quien desempeña funciones en el Programa de Genética Humana, ICBM, Facultad de Medicina, Universidad de Chile.

El Comité revisó los siguientes documentos del estudio:

- *Proyecto de Investigación in extenso*
- *Consentimiento informado*
- *CV del Investigador responsable y de los Co-investigadores*
- *Carta compromiso del investigador para comunicar los resultados del estudio una vez finalizado éste.*

El proyecto y los documentos señalados en el párrafo precedente han sido analizados a la luz de los postulados de la Declaración de Helsinki, de las Pautas Éticas Internacionales para la Investigación Biomédica en Seres Humanos CIOMS 2002, y de las Guías de Buena Práctica Clínica de ICH 1996.

Teléfono: 9786923 Fax: 9786189 Email: ceiha@med.uchile.cl



UNIVERSIDAD DE CHILE
FACULTAD DE MEDICINA
COMITÉ DE ÉTICA DE INVESTIGACIÓN EN SERES HUMANOS



31 MAYO 2011

Sobre la base de esta información el Comité de Ética de la Investigación en Seres Humanos de la Facultad de Medicina de la Universidad de Chile se ha pronunciado de la siguiente manera sobre los aspectos del proyecto que a continuación se señalan:

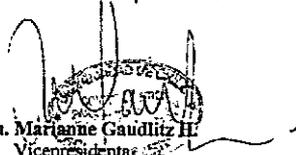
- a) Carácter de la población estudiada: Investigación no terapéutica, población no cautiva.
- b) Utilidad del Proyecto: sí
- c) Riesgos y Beneficios: Riesgos no hay.
- d) Protección de los participantes: Está bien asegurado en el consentimiento informado.
- e) Notificación oportuna de reacciones adversas: no aplica.
- f) El investigador responsable se ha comprometido a entregar los resultados del estudio a este Comité al finalizar el proyecto.

Por lo tanto, el comité estima que el estudio propuesto está bien justificado y que no significa para los sujetos involucrados riesgos físicos, psíquicos o sociales mayores que mínimos.

Este comité también analizó y aprobó el correspondiente documento de Consentimiento Informado en su versión original del 30 de mayo de 2011, que se adjunta firmado, fechado y timbrado por el CEISH.

En virtud de las consideraciones anteriores el Comité otorga la aprobación ética para la realización del estudio propuesto, dentro de las especificaciones del protocolo.

Santiago, 31 de mayo de 2011.


Sra. Marianne Gaudlitz H.
Vicepresidenta
Comité de Ética en Investigación
en Seres Humanos

MGH/mva.
c.c.: Archivo Proy. 023-2011

☛ **Teléfono: 9786923 Fax: 9786189 Email: ceiha@med.uchile.cl**

7. BIBLIOGRAFÍA

- Antoun, A., Jobson, S., Cook, M., O'Callaghan, C., Moss, P., Briggs, D. (2010). Single nucleotide polymorphism analysis of the NKG2D ligand cluster on the long arm of chromosome 6: Extensive polymorphisms and evidence of diversity between human populations. *Human Immunology*. **71**: 610-620.
- Asaka, M., Kato, M., Sakamoto, N. (2014). Roadmap to eliminate gastric cancer with *Helicobacter pylori* eradication and consecutive surveillance in Japan. *Journal of Gastroenterology*. **49**: 1-8.
- Bahram, S., Bresnahan, M., Geraghty, D. (1994). A second lineage of mammalian major histocompatibility complex class I genes. *Proceedings of the National Academy of Sciences of the United States of America*. **91**: 6259-6263.
- Bauer, S., Groh, V., Wu, J., Steinle, A., Phillips, J., Lanier, L. (1999). Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA. *Science*. **285**: 727-729.
- Bronner, I., Quail, M., Turner, D., Swerdlow, H. (2013). Improved Protocols for Illumina Sequencing. *Current protocols in Human Genetics*. **79**: 1-42.
- Burnet, F. M. (1971). Immunological Surveillance in Neoplasia. *Immunological Reviews*. **7**: 3-25.
- Caligiuri, M. (2008). Human natural killer cells. *Blood*. **112**: 461-469.
- Campisi, J. (2013). Aging, cellular senescence, and cancer. *Annual Review of Physiology*. **75**: 685-705.
- Cerwenka, A., & Lanier, L. (2001). Natural killer cells, viruses and cancer. *Nature Reviews Immunology*. **1**: 41-49.
- Chalupny, N., Sutherland, C., Lawrence, W., Rein-Weston, A., Cosman, D. (2003). ULBP4 is a novel ligand for human NKG2D. *Biochemical Biophysical Research Communications*. **305**: 129-135.
- Chan, C., Crafton, E., Fan, H., Yoshimura, K., Skarica, M., Lanier, L., Pardoll, D., Housseau, F. (2006). Interferon-producing killer dendritic cells provide a link between innate and adaptive immunity. *Nature Medicine*. **12**: 207-213.
- Chang, C., Dietrich, J., Harpur, A., Lindquist, J., Haude, A., Loke, Y., King, A., Colonna M., Trowsdale J., Wilson, M. (1999). Cutting edge: KAP10, a novel

- transmembrane adapter protein genetically linked to DAP12 but with unique signaling properties. *Journal of Immunology*. **163**: 4651-4654.
- Chen, D. & Gyllensten, U. (2014). MICA polymorphism: biology and importance in cancer. *Carcinogenesis*. **35**: 2633-2642.
- Choy, M. & Phipps, M. (2010). MICA polymorphism: biology and importance in immunity and disease. *Trends of Molecular Medicine*. **16**: 97-106.
- Cooper, D. (2010). Functional intronic polymorphisms: Buried treasure awaiting discovery within our genes. *Human Genomics*. **4**: 284-288.
- Cosman, D., Mullberg, J., Sutherland, C., Chin, W., Armitage, R., Chalupny, N. (2001). ULBPs, novel MHC class I-related molecules, bind to CMV glycoprotein UL16 and stimulate NK cytotoxicity through the NKG2D receptor. *Immunity*. **14**: 123-133.
- Coudert, J. & Held, W. (2006). The role of the NKG2D receptor for tumor immunity. *Seminars Cancer Biology*. **16**: 333-343.
- De Baets, G., Van Durme, J., Reumers, J., Maurer-Stroh, S., Peter Vanhee, Dopazo, J., Schymkowitz, D. (2012). SNPeffect 4.0: on-line prediction of molecular and structural effects of protein-coding variants. *Nucleic Acids Research*. **40**: 935-939.
- Deniz, G., Erten, G., Kocacik, D., Aktas, E., Akdis, C., Akdis, M. (2008). Regulatory NK cells suppress antigen-specific T cell responses. *Journal of Immunology*. **180**: 850-857.
- Dennert, G., Anderson, C., Prochazka, G. (1987). High activity of N-alpha-benzyloxycarbonyl-L-lysine thiobenzyl ester serine esterase and cytolytic perforin in cloned cell lines is not demonstrable in in-vivo-induced cytotoxic effector cells. *Proceedings of the National Academy of Sciences of the United States of America*. **84**: 5004-5008.
- Dicken, B., Bigam, D., Cass, C., Mackey, J., Joy, A., Hamilton, S. (2005). Gastric adenocarcinoma: review and considerations for future directions. *Annals of Surgery*. **241**: 27-39.
- Dunn, G., Old, L., Schreiber, R. (2004). The three Es of cancer immunoediting. *Annual Review of Immunology*. **22**: 329-360.
- Dunn, G., Koebel, C., Schreiber, R. (2006). Interferons, immunity and cancer immunoediting. *Nature Reviews Immunology*. **6**: 836-848.

- Eagle, R. & Trowsdale, J. (2007). Promiscuity and the single receptor: NKG2D. *Nature Reviews Immunology*. **7**: 737-744.
- Eagle, R., Flack, G., Warford, A., Jafferji, I., Boyle, L., Barrow, A., Trowsdale, J. (2009) Cellular expression, trafficking, and function of two isoforms of human ULBP5/RAET1G. *PLoS ONE*. **4**: e4503.
- Ferlazzo, G., Thomas, D., Lin, S., Goodman, K. (2004). The abundant NK cells in human secondary lymphoid tissues require activation to express killer cell Ig-like receptors and become cytolytic. *Journal of Immunology*. **172**: 1455-1462.
- Frigoul, A. & Lefranc, M. (2005). MICA: Standardized IMGT allele nomenclature, polymorphisms and diseases. *Recent Research Developments in Human Genetics*. **3**: 95-145.
- Goldstein, D. & Cavalleri, G. (2005). Genomics: Understanding human diversity. *Nature*. **437**: 1241-1242.
- Gonzalez, S., Groh, V., Spies, T. (2006). Immunobiology of human NKG2D and its ligands. *Current Topics in Microbiology and Immunology*. **298**: 121-138.
- Gregoire, C., Chasson, L., Luci, C., Tomasello, E., Geissmann, F., Vivier, E., Waizer, T. (2007). The trafficking of natural killer cells. *Immunological Reviews*. **220**, 169-182.
- Groh, V., Bahram, S., Bauer, S., Herman, A., Beauchamp, M., Spies, T. (1996). Cell stress-regulated human major histocompatibility complex class I gene expressed in gastrointestinal epithelium. *Proceedings of the National Academy of Sciences of the United States of America*. **93**: 12445-12450.
- Groh, V., Wu, J., Yee, C., Spies, T. (2002). Tumour-derived soluble MIC ligands impair expression of NKG2D and T-cell activation. *Nature*. **419**: 734-738.
- Hanahan, D. & Weinberg, R. (2011). Hallmarks of cancer: the next generation. *Cell*. **144**: 646-674.
- Harismendy, O., Schwab, R., Bao, L., Olson, J., Rozenzhak, S., Kotsopoulos, S., Frazer, K. (2011). Detection of low prevalence somatic mutations in solid tumors with ultra-deep targeted sequencing. *Genome Biology*. **12**: 124.
- Houchins, J., Yabe, T., McSherry C., Bach, F. (1991). DNA sequence analysis of NKG2D, a family of related cDNA clones encoding type II integral membrane proteins on human natural killer cells. *The Journal of Experimental Medicine*. **173**: 1017-20.

- Illumina Inc. (2013). Análisis de datos mediante el uso del flujo de trabajo Resequencing (Resecuenciación)
- Illumina Inc. (2015). MiSeq System. System Specification Sheet: Sequencing
- Itriago, L., Silva, N., Cortes, G. (2013) Cáncer en Chile y el Mundo: Una Mirada Epidemiológica, Presente y Futuro. *Revista Médica Clínica Las Condes*. **24**: 531-552.
- Kandulski, A., Malfertheiner, P., Wex, T. (2010). Role of regulatory T-cells in H. pylori- induced gastritis and gastric cancer. *Anticancer Research*. **30**: 1093-1103.
- Karre, K., Ljunggren, H., Piontek, G., Kiessling, R. (2005). Selective rejection of H-2-deficient lymphoma variants suggests alternative immune defence strategy. (1986). *Journal of Immunology*. **174**: 6566-6569.
- Kassahn, K., Holmes, O., Nones, K., Patch, A., Miller, D., Christ, A. (2013) Somatic Point Mutation Calling in Low Cellularity Tumors. *PLoS ONE*. **8**: e74380.
- Kennedy, S., Loeb, L., Herr, A. (2012). Somatic Mutations in Aging, Cancer and Neurodegeneration. *Mechanisms of Ageing and Development*. **133**: 118–126.
- Kumar, P., Henikoff, S., Ng, P. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*. **4**:1073-1081.
- Kumar, V., Yi Lo P., Sawai, H., Kato, N., Takahashi, A., Deng, Z. (2012). Soluble MICA and a *MICA* variation as possible prognostic biomarkers for HBV-induced hepatocellular carcinoma. *PLoS ONE*. **7**: e44743.
- Le Clerc, S., Delaneau, O., Coulonges, C., Labib, T., Noirel J., Zagury, J. (2014). Evidence after imputation for a role of MICA variants in nonprogression and elite control of HIV type 1 infection. *The Journal of Infectious Diseases*. **210**: 1946-1950.
- Lee, K., Caceres, D., Varela, N., Csendes, D., Ríos, R., Quiñones, S. (2006). Allelic variants of cytochrome P4501A1 (CYP1A1), glutathione S transferase M1 (GSTM1) polymorphisms and their association with smoking and alcohol consumption as gastric cancer susceptibility biomarkers. *Revista Médica de Chile*. **134**: 1107-1115.
- Loza, M., Zamai, L., Azzoni, L., Rosati, E., Perussia, B. (2002). Expression of type 1 (interferon gamma) and type 2 (interleukin-13, interleukin-5) cytokines at

- distinct stages of natural killer cell differentiation from progenitor cells. *Blood*. **99**: 1273-1281.
- McLean, M. & El-Omar, E. (2014). Genetics of gastric cancer. *Nature reviews. Gastroenterology & Hepatology*. **11**: 664-674.
- Medina, E. & Kaempffer, A. (2000). Adult mortality in Chile. *Revista Médica de Chile*. **128**: 1144-1149.
- Minoche, A., Dohm, J., Himmelbauer, H. (2011). Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems. *Genome Biology*. **12**: 112.
- MINSAL, Ministerio de Salud, República de Chile (2010). *Guía Clínica para Cáncer Gástrico*.
- Mizuki, N., Ota, M., Kimura, M., Ohno, S., Ando, H., Yamazaki, M., Nakamura, S., Bahram, S. (1997). Triplet repeat polymorphism in the transmembrane region of the MICA gene: A strong association of six GCT repetitions with Behcet disease. *Proceedings of the Natural Academy of Science of the USA*. **94**: 1298-1303.
- Nausch N. (2008). NKG2D ligands in tumor immunity. *Oncogene*. **27**: 5944-5958.
- Obuchi, N., Takahashi, M., Satoh, M., Arimura, T., Akai, J., Naruse, T., Inoko, H., Numano, F., Kimura, A. (2001). Identification of MICA alleles with a long Leu-repeat in the transmembrane region and no cytoplasmic tail due to a frameshift-deletion in exon 4. *Tissue Antigens*. **57**: 520-535.
- Okines, A., Verheij, M., Allum, W., Cunningham, D., Cervantes, A. (2010). Gastric cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*. **5**: 50-54.
- Pérez-Rodríguez, M., Corell, A., Argelo, J., Cox, S., Marsh, S., Madrigal, J. (2000). A new MICA allele with ten alanine residues in the exon 5 microsatellite. *Tissue Antigens*. **55**: 162-165.
- Podack, E. & Dennert, G. (1983). Assembly of two types of tubules with putative cytolytic function by cloned natural killer cells. *Nature*. **302**: 442-445.
- Robinson, J., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E., Getz, G., Mesirov, J. (2011). Integrative Genomics Viewer. *Nature Biotechnology*. **29**: 24-26.

- Rutvisuttinunt, W., Chinnawirotpisan, P., Simasathien, S., Shrestha, S., Yoon, I., Klungthong, C. (2013). Simultaneous and complete genome sequencing of influenza A and B with high coverage by Illumina MiSeq Platform. *Journal of Virological Methods*. **193**: 394–404.
- Salih, H., Rammensee, H., Steinle, A. (2002). Cuttingedge: down-regulation of MICA on human tumors by proteolytic shedding. *The Journal of Immunology*. **169**: 4098-4102.
- Shifrin, N., Raulet, D., Ardolino, M. (2014). NK cell self tolerance, responsiveness and missing self recognition. *Seminars in Immunology*. **26**: 138-144.
- Smyth, M., Swann, J., Cretney, E., Zerafa, N., Yokoyama, W., Hayakawa, Y. (2005). NKG2D function protects the host from tumor initiation. *The Journal of Experimental Medicine*. **202**: 583-588.
- Street, T., Bradley, C., Barrick, D. (2005). An improved experimental system for determining small folding entropy changes resulting from proline to alanine substitutions. *Protein Science*. **14**: 2429–2435.
- Tamaki, S., Sanefuzi, N., Ohgi, K., Imai, Y., Kawakami, M., Yamamoto, K., Ishitani, A., Hatake, K., Kirita, T. (2007). An association between the MICA-A5.1 allele and an increased susceptibility to oral squamous cell carcinoma in Japanese patients. *Journal of Oral Pathology and Medicine*. **36**: 351-356.
- Tamaki, S., Kawakami, M., Yamanaka, Y., Shimomura H., Ishida, J., Yamamoto, K., Ishitani, A., Hatake, K., Kirita, T. (2009). Relationship between soluble MICA and the MICA A5.1 homozygous genotype in patients with oral squamous cell carcinoma. *Clinical Immunology*. **130**: 331-337.
- Tian, D., Wang, Q., Zhang, P., Araki, H., Yang, S., Kreitman, M., Nagylaki, T., Hudson, R., Bergelson, J., Chen, J. (2008). Single-nucleotide mutation rate increases close to insertions/deletions in eukaryotes. *Nature*. **455**: 105–108.
- Tian, W., Boggs, D., Ding, W., Chen, D., Fraser, P. (2001) MICA genetic polymorphism and linkage disequilibrium with HLA-B in 29 African-American families. *Immunogenetics*. **53**: 724–728. |
- Torre, L., Bray, F., Siegel, R., Ferlay, J., Lortet, J., Jemal, A. (2015). Global cancer statistics, 2012. *CA: A Cancer Journal for Clinicians*. **65**: 87-108.
- Trapani, J. & Smyth, M. (2002). Functional significance of the perforin/granzyme cell death pathway. *Nature Reviews. Immunology*. **2**: 735-747.

- Trinchieri, G. (1989). Biology of natural killer cells. *Advances in Immunology*. **47**: 187-376.
- Trivedi, U., Cézard, T., Bridgett, S., Montazam, A., Nichols, J., Blaxter, M., Gharbi, K. (2014). Quality control of next-generation sequencing data without a reference. *Frontiers in Genetics*. **5**: 111.
- Vega, F., Brown, S., Chen, H., Shen, E., Nair, M., Ceja-Navarro, J. (2015). Draft genome of the most devastating insect pest of coffee worldwide: the coffee berry borer, *Hypothenemus hampei*. *Scientific Reports*. **5**: 125.
- Vetter, C., Lieb, W., Brocker, E., Becker, J. (2004). Loss of nonclassical MHC molecules MIC-A/B expression during progression of uveal melanoma. *British Journal of Cancer*. **91**: 1495–1499.
- Visser, K., Eichten, A., Coussens, L. (2006). Paradoxical roles of the immune system during cancer development. *Nature reviews. Cancer*. **6**: 24-37.
- Wang, Z. & Burge, C. (2008). Splicing regulation: From a parts list of regulatory elements to an integrated splicing code. *RNA*. **14**: 802–813.
- Wu, J., Song, Y., Bakker, A., Bauer, S., Spies, T., Lanier, L., Phillips J. (1999). An activating immunoreceptor complex formed by NKG2D and DAP10. *Science*. **285**: 730-732.
- Wu, J., Cherwinski, H., Spies, T., Phillips, J., Lanier, L. (2000). DAP10 and DAP12 form distinct, but functionally cooperative, receptor complexes in natural killer cells. *The Journal of Experimental Medicine*. **192**: 1059-68.
- Yost, S., Smith, E., Schwab, R., Bao, L., Jung, H., Wang, X., Frazer, K. (2012). Identification of high-confidence somatic mutations in whole genome sequence of formalin-fixed breast cancer specimens. *Nucleic Acids Research*. **40**: 107.
- Yost, S., Alakus, H., Matsui, H., Schwab, R., Jepsen, K., Frazer, K., Harismendy, O. (2013). Mutoscope: sensitive detection of somatic mutations from deep amplicon sequencing. *Bioinformatics*. **29**: 1908–1909.
- Zamai, L., Ahmad, M., Bennett, I., Azzoni, L. (1998). Natural killer (NK) cell- mediated cytotoxicity: differential use of TRAIL and Fas ligand by immature and mature primary human NK cells. *The Journal of Experimental Medicine*. **188**: 2375-80.
- Zhang, Q., Zhang, J., Jin, H., Sheng, S. (2013). Whole transcriptome sequencing identifies tumor-specific mutations in human oral squamous cell carcinoma. *BMC Medical Genomics*. **6**: 28.

Zwirner, N., Fernandez-Vina, M., Stastny, P. (1998). MICA, a new polymorphic HLA-related antigen, is expressed mainly by keratinocytes, endothelial cells, and monocytes. *Immunogenetics*. 47: 139-148.