UNIVERSIDAD DE CHILE
Facultad de Ciencias
Escuela de Postgrado

"VALIDACIÓN Y EVALUACIÓN DE ALGORITMOS PARA DETECCIÓN DE *INDELS* EN EL GEN *MICA* EN CÁNCER GÁSTRICO"

Tesis entregada a la Universidad de Chile en cumplimiento parcial de los requisitos para optar al grado de Magíster en Ciencias Biológicas.

Por

VALENTINA CONSTANZA GÁRATE CALDERÓN

Enero, 2018

Directora de Tesis:
Dra. María Carmen Molina Sampayo

Co-Director de Tesis:
Dr. Ricardo Amado Armisén Yañez

Santiago – Chile

FACULTAD DE CIENCIAS

UNIVERSIDAD DE CHILE

INFORME DE APROBACION

TESIS DE MAGÍSTER

Se informa a la Escuela de Postgrado de la Facultad de Ciencias que la Tesis de Doctorado presentada por la candidata.

VALENTINA CONSTANZA GÁRATE CALDERÓN

Ha sido aprobada por la comisión de Evaluación de la tesis como requisito para optar al grado de Magíster en Ciencias Biológicas, en el examen de Defensa Privada de Tesis rendido el día 18 de Enero De 2018

Director de Tesis:
Dra. María carmen Molina S. .....

Co-Director de Tesis
Dr. Ricardo Armisén Y. .....

Comisión de Evaluación de la Tesis

Dr. Mauricio González .....

Dr. Christian Hodar .....

A mis padres, Mariela y Pablo,

Sin ustedes esto no habría sido posible.

AGRADECIMIENTOS

En primer lugar, quiero agradecer a la Dra. María Carmen Molina y en especial al Dr. Ricardo Armisén 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). A los chicos del IBT, Norbert, Mati, Basti y Alfons, por su constante apoyo.

Gracias a todos mis colegas CEMP, que me muestran día a día que con una mentalidad desafiante se puede llegar lejos. En especial agradezco a Germán y Gonzalo, por guiarme con paciencia por el mundo de la bioinformática.

Agradezco a mis compañeros Biotecs, porque han estado en muchos momentos dándome fuerzas para completar este desafío y hemos construido lindos recuerdos de esta aventura.

A mis Fresscas de la vida: Maca, Pazi, Anto y Fran 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, ubicado en el instituto de Ciencias Biomédicas (ICBM) de la facultad de Medicina de la Universidad de Chile; en colaboracion con Centro de Excelencia en Medicina de Precisión (CEMP) de Pfizer Chile. Su realización fue posible con el financiamiento de los proyectos FONDECYT 1130330 (M.C.M), y 1151446 (R.A), FONDEF D11I1029 (R.A) y ANILLO ACT1115 (R.A).

ÍNDICE

LISTA DE TABLAS

LISTA DE FIGURAS

LISTA DE ABREVIATURAS

ADN	Ácido desoxirribonucleico
ARN	Ácido ribonucleico
CG	Cáncer gástrico
dNTP	Desoxinucleótido trifosfato
IGV	Visualizador integrado de genomas
IMGT	Base de datos ImmunoGenetics
Indel	Inserción/Deleción
MHC-I	Complejo principal de histocompatibilidad de clase I
MICA	Proteína relacionada con la cadena A 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
Pb	Pares de bases
PCR	Reacción en cadena de la polimerasa
SIFT	Sorting intolerant from tolerant
sMICA	MICA soluble
SNP	Polimorfismo de un sólo nucleótido
STR	Repetición corta en tándem, microsatélite
SNV	Variante de un sólo nucleótido
SVC	Somatic Variant Caller
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 anti-tumoral y uno de sus componentes más importantes son las células Natural Killer (NK), las cuales están especializas 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; NKG2D es le receptor de activación mejor caracterizado, y sus ligandos (NKG2DL), incluyendo la proteína MICA interactuan con el receptor NKG2D e inducen la secreción de citoquinas y/o citotoxicidad mediada por células NK hacia las células diana, proceso clave en la inmunovigilancia tumoral. El ligando MICA está ausente en la mayoría de las células, pero su expresión puede ser inducida por infecciones y transformación oncogénica. Se han identificado Inserciones y Deleciones (*Indels*) en la región del gen *MICA* que codifica el dominio transmembrana (TM) de la proteína, donde la presencia de repetidos cortos en tándem (STR) en el exón 5' genera una serie de variantes, y, en particular, A5.1STR resulta en un codón de stop prematuro codificando para una proteína trunca. En este punto, se ha reportado que pacientes con carcinoma hepatocelular con el genotipo homocigoto A5.1 presentan niveles más altos de MICA solubles y una menor tasa de supervivencia.

Debido a la importancia que tendrían estas mutaciones en la funcionalidad del ligando, en esta tesis se plantea la utilización de diversos algoritmos de alineamiento local para la detección de *Indels* en la secuencia del gen *MICA*, los que podrían estar implicados en la evasión inmune en cáncer gástrico. El objetivo de este trabajo fue identificar estas mutaciones en datos de pacientes chilenos con cáncer gástrico y analizar el posible efecto 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 y se realizó secuenciación masiva para el gen *MICA* en la plataforma MiSeq de Illumina. El análisis bioinformático de las secuencias se realizó con los algoritmos SVC (Illumina), Pindel, Scalpel, SOAPindel y Varscan. Como resultado, se identificaron 5

coordenadas cromosómicas que fueron comunes entre los 5 algoritmos utilizados. A continuación, se validó la detección de *Indels* por secuenciacion masiva mediante análisis de secuenciación por Sanger, y se determinó que los algoritmos menos restrictivos fueron SOAP y Pindel, mientras que Scalpel y SVC se posicionan como los programa con los resultados más cercanos a lo obtenido por el método de Sanger.

En el presente estudio, se identificaron diversos *Indels* en el dominio TM de MICA, destacando a aquellos presentes en A5.1 (codifica para MICA trunca) y A9 (se adicionan 4 aminoacidos a la secuencia de MICA) como los más frecuentes. La variante MICA-A9 estaría presente en 40% de las muestras analizadas, lo que coincidiría con un mayor riesgo de cáncer gástrico, tal como descrito en pacientes Taiwaneses, y mayor riesgo de desarrollar carcinoma hepatocelular inducido por virus de hepatitis B. Complementariamente, los resultados entregados por el programa SIFT indican que, si bien las variantes detectados no tienen un efecto directo en el sitio de unión con el receptor NKG2D, la modificación de la región TM podría estar dando cuenta de la liberación de ligandos solubles o alteraciones estructurales de este dominio por la adición de aminoácidos en su secuencia. Nuestras observaciones sugieren que la presencia de *Indels* en el gen *MICA*, en pacientes con adenocarcinoma gástrico, podría corresponder a una estrategia de evasión inmune que esté favoreciendo el establecimiento y desarrollo tumoral.

ABSTRACT

Gastric cancer (GC) is one of the leading causes of cancer mortality in Chile. Innate immune response is the first line of host defense against tumors, and one of its most important components are the Natural Killer (NK) cells, which are specialized in the elimination of virus-infected cells and tumor cells. A complex balance of inhibiting and activating receptors regulates the activity of NK cells. NKG2D is the best characterized activating receptor, whose ligands (NKG2DL), including MICA protein, interact with NKG2D receptor and trigger cytotoxicity mediated by NK cells toward target cells, a key process in the immune surveillance against cancer. MICA is absent in the majority of cells, but its expression can be induced by infection and oncogenic transformation. Several lines of evidence have identified insertions and deletions (Indels) in the transmembrane domain (TM) of MICA, where the insertion of short tandem repeat (STR) in exon 5 generates a series of alleles, in particular the A5.1STR allele results in a premature stop codon, coding for a truncated protein. Indeed, patients with hepatocellular carcinoma bearing the homozygous genotype A5.1 present higher levels of soluble MICA and lower survival rates.

Since these mutations may be important to NKG2DL functionality, in this thesis we used a group of local alignment algorithms to detect Indels in *MICA*, especially those that can be involved in immune evasion strategies in gastric cancer. The aim of this work was to identify mutations in *MICA* gene in Chilean patients with gastric cancer and analyze the structural changes these mutations would cause in the protein. Fifty patients with gastric adenocarcinoma were included in this study, and all of them were reluctant from the Department of Digestive Surgery of the Hospital del Salvador, University of Chile. Mutations in the sequence of MICA gene were detected by massive sequencing using the MiSeq platform of Illumina. The bioinformatic analysis of the gene sequences was performed with the SVC (Illumina), SOAPindel, Pindel, Scalpel and Varscan algorithms, which allowed the identification of 5 common chromosomal coordinates. We checked the detection of Indels by NGS through analysis of Sanger sequencing and

determined that the less restrictive algorithms were SOAP and Pindel, while Scalpel and SVC showed the closest results to those obtained by the Sanger method.

In this study, we identified several Indels in the TM domain of MICA, particularly those coding for MICA A5.1 (truncated MICA) and A9 (4 amino acids are added to the sequence of MICA) as the most frequent Indels. The MICA-A9 allele was present in 40% of the analyzed samples, which may be related with an increased risk for gastric cancer, as described in Taiwanese patients and higher risk of developing hepatocellular carcinoma induced by hepatitis B virus. In addition, the results obtained using the SIFT program indicate that, even though the variants detected here do not have a direct effect on the binding site with NKG2D, the modifications observed in the TM region of MICA may result in the release of soluble ligands or structural alterations of this domain by the addition of amino acids in its sequence.

Our results suggest that the presence of Indels in the *MICA* gene, in patients with gastric adenocarcinoma, may play a role in an immune evasion strategy, thus favoring tumor establishment and development.

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 al año) 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 & El-Omar, 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; Lamb & Chen, 2013)

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 publicas 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 órganos, células y moléculas 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 ocacionar 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 activación y tolerancia, los que están balanceados por complejos mecanismos de control que favorecen su acción efectora. 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 (Diefenbach & Raulet, 2002; 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 medula ósea, y la mayoría se encuentra localizada en la sangre periférica, linfonodos, bazo y médula ósea y en humanos, comprenden entre un 5-20% de los linfocitos de sangre periférica (Ferlazzo *y col.*, 2004).

Las células NK eliminan las células blanco (células tumorales, envejecidas 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, las cuales promueven lisis celular por osmosis e inducen apoptosis de las

células blanco, respectivamente (Podack & Dennert, 1983; Dennert *y col.*, 1987; Trapani & Smyth, 2002).

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 inmunoreguladoras, 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 inhibitorios y activadores, los que se unen a moléculas presentes en la superficie de células blanco tales como MHC-I (Complejo principal de histocompatibilidad de clase I) y moléculas relacionadas con MHC-I, respectivamente. La activación de las células blanco 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 blanco (Smyth *y col.*, 2005). En este sentido, las células blanco aumentan la expresión de ligandos del receptor de activación de células NK frente a la infección, envejecidas o transformadas. Por otro lado, también pueden disminuir la expresión de la molécula MHC-I en la superficie de la célula blanco, ya sea por mutaciones genéticas que ocurren durante el desarrollo tumoral o por mecanismos de evasión viral. Esto último, 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). Así, el resultado de estas interacciones va a determinar si la célula NK es finalmente activada, si se producen citoquinas y/o si la célula blanco es eliminada (Caligiuri, 2008).

1.4 Receptor NKG2D y sus ligandos

El receptor NKG2D (Natural Killer Group 2, type D) es uno de los receptores de activación mejor caracterizado, se expresa en todas las células NK y ha sido identificado en linfocitos Tαβ CD8+, T γδ 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 transmembrana de tipo II unidas por un puente de disulfuro. El dominio de transmembrana 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 α1, α2 y α3, pero se diferencian de éstos al no unir β2- microglobulina ni presentar antígenos (Zwirner *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 α1 y α2 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 servirían para poder detectar las células transformadas (Nausch, 2008; Fernández-Messina *y col.*, 2012). 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 Inmunoedición 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.

La evasión 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), entre otros.

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 expresando estas moléculas. Teniendo en cuenta esta contradicción, se ha planteado que podría existir una selección negativa sobre las células tumorales que expresan NKG2DL en su superficie. Por un lado, en investigaciones en melanoma se documenta 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; Marcus *y col.*, 2014). 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 (Coudert *y col.*, 2008; Champsaur & Lanier, 2010) . 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 & Held, 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 variantes polimórficas 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).

En cáncer gástrico hemos informado que NKG2DL está presente en el tumor primario CG (Ribeiro *y col.*, 2016) y en líneas celulares CG (Garrido-Tapia *y col.*, 2017), al igual que en melanoma (Serrano *y col.*, 2011). Las células tumorales primarias de los pacientes con CG tienen una mayor expresión de MICA / B, mientras que sus linfocitos expresaron niveles más bajos de NKG2D que la mucosa gástrica circundante (Ribeiro *y col.*, 2016). Hemos demostrado que la expresión de MICA en los tumores también se correlaciona con los parámetros clínico-patológicos de la enfermedad, ya que los pacientes con tumores mayores a 5 cm de diámetro que expresan MICA tienen menores tasas de supervivencia que aquellos que no expresan MICA (Ribeiro *y col.*, 2016). Además, también hemos encontrado que sMICB, pero no sMICA, disminuye la expresión de NKG2D en las células NK (Garrido-Tapia *y col.*, 2017), en contraste con los informes de otros autores, lo que podría deberse a las variaciones polimórficas en este ligando. Por lo tanto, es incorrecto considerar a todos los alelos de NKG2DL como moléculas biológicamente equivalentes.

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).

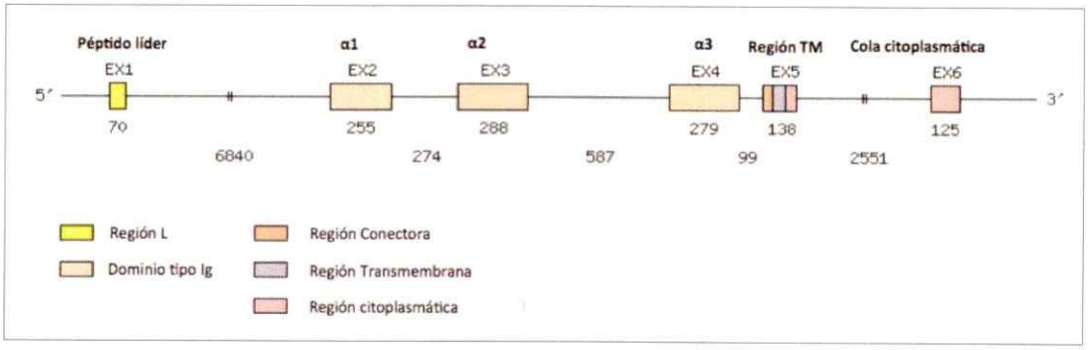


Figura 1. Organización exón-intrón del gen *MICA*. El gen *MICA* contiene 6 exones que codifican para un péptido líder (Región L, Exón 1), 3 dominios extracelulares (α1-3) (Exones 2,3 y 4, respetivamente), un dominio de transmembrana (Exón 5) y un dominio citoplasmático (Exón 6). El largo de los exones 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 105 alelos para este gen, codificando 82 variantes de la proteína (base de datos IMGT, http://www.imgt.org). Donde la mayoría de los sitios polimórficos varían en posiciones nucleotídicas no redundantes, con predominancia en los dominios α2 y α3 implicados tanto en la interacción con NKG2D (Figura 1) así como también con el sitio de clivaje por metaloproteasas (Salih *y col.*, 2002).

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 costos 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 (Meldrum *y col.*, 2011), permitiendo 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.

En este contexto, dos estudios de asociación del genoma completo (GWAS) han identificado un *loci* en *MICA* que estarían relacionados con susceptibilidad a carcinoma hepatocelular inducido por el virus de la hepatitis B y a neoplasia cervical, respectivamente (Tong *y col.*, 2013; Chen & Gyllensten, 2014).

Las mutaciones de tipo Inserción/Deleción (*Indels*) son el segundo tipo más común de variación genética en el genoma humano, afectando a una multitud de rasgos humanos y enfermedades. Desde un punto funcional, se han identificado inserciones en el dominio transmembrana (TM) del ligado MICA, donde la inserción de repetidos cortos en tándem (STR) en el exón 5 resultan en una serie de variantes, y en particular la inserción de un nucleótido rs67841474 (GCT > GGCT) en la región TM de las variantes A5.1STR genera un desplazamiento en el marco abierto de lectura, lo que resulta en un codón de stop prematuro, codificando para una proteína trunca (Tamaki *y col.*, 2007). Además, pacientes con carcinoma hepatocelular con el genotipo homocigoto A5.1 han reportado

niveles más altos de MICA solubles y una menor tasa de supervivencia (Kumar *y col.*, 2012).

En tanto, de las diversas variantes descritas para el exón 4 del gen MICA, el impacto funcional más dramático se observa para la deleción rs199503730, con un efecto más severo incluso que para mutaciones no sinónimas cercanas, debido a que la deleción induce un desplazamiento del marco de lectura justo antes de la sección TM de la proteína MICA, conllevando una variación en su secuencia que afecta directamente a la región hidrofóbica de la región TM (Le Clerc *y col.*, 2014).

1.7 Desafíos de la bioinformática en la búsqueda de *Indels* para NGS.

Se ha analizado la asociación entre *Indels* y sustituciones de única base (SNP) en la base de datos Catálogo de Mutaciones Somáticas en Cáncer (COSMIC), y se ha evidenciado una fuerte correlación entre *Indels* y SNP en genes relacionados con el cáncer, demostrando que tienden a concentrarse en el mismo locus de secuencias codificantes dentro de las mismas muestras de pacientes (Tian *y col.*, 2008). Además se han identificado menos mutaciones de desplazamiento de lectura en oncogenes, en comparación con genes supresores de tumores, debido a su diferente función en la oncogénesis; donde tal situación estaría dando cuenta que *Indels* pueden ser las "mutaciones conductoras" en la oncogénesis. Estos resultados contribuyen a una nueva comprensión de los patrones de mutaciones y la relación entre *Indels* y el cáncer (Iengar, 2012; Yang *y col.*, 2015).

Recientemente, se encontró que la tasa de sustitución de nucleótidos (SNP) es significativamente elevada alrededor de *Indels* y se correlaciona tanto con el tamaño y abundancia del *Indel* (Tian *y col.*, 2008), lo que sugiere un rol importante de los *Indels* en las mutaciones de los genes relacionados con el cáncer.

La mayoría de los programas bioinformáticos para la detección de mutaciones en datos de NGS, alinean un *read* (lectura) al genoma de referencia a la vez y luego analizan el alineamiento con el objetivo de detectar alguna mutación (DePristo *y col*, 2011). Si bien este enfoque funciona de manera óptima en la identificación de mutaciones simples de tipo SNV, donde *reads* con bases mutadas son alineados al genoma de referencia a lo largo de la mutación; para el análisis de *Indels*, este proceso se vuelve cada vez menos efectivo. Esto esta dado por el hecho que *reads* que presentan inserciones o deleciones en su secuencia generan complicaciones en el alineamiento ya que los extremos del *read* podrían no presentar bases que alineen a la referencia, obligando a recortar los *reads* (*trimming*).

En la actualidad existen dos enfoques bioinformaticos para la identificación de *Indels* desde datos de NGS. Por un lado se tienen los algoritmos de tipo *Breakpoint* (puntos de quiebre), los que se centran en identificar secuencias en los extremos de los *reads* que no mapean al genoma de referencia y las cortan en pequeños fragmentos de tamaño definido (k-mer), realizando un ensamblaje *de novo* de estos fragmentos para generar finalmente una secuencia concenso (contig) a partir de *reads* realineados, la que estaría dando cuenta de la secuencia de los *Indels* en estudio (Korbel *y col.*, 2007; Tattini *y col.*, 2015). Se ha descrito que este tipo de algoritmo permite la identificación de *Indels* de

tamaño 2 a 10.000 pb, incluyendo *Indels* complejos y variantes estructurales, por lo que se denominan algoritmos de amplio espectro (Ratan *y col.*, 2015).

El otro enfoque lo componen algoritmos de tipo Soft-clip (recorte suave), los que se centran en identificar los extremos de los *reads* que no mapean al genoma de referencia, y realinean estas secuencias de forma restringida a los segmentos adyacentes en los extremos de los *reads*, permitiendo la identificación de *Indels* en la proximidad de la región analizada (Schröder *y col.*, 2014). Con este tipo de análisis estos algoritmos logran identificar *Indels* de largo menor a 400 pb, permitiendo analizar secuencia acotadas con alta precisión (Narzisi & Schatz, 2015).

Cabe señalar que si bien algoritmos desarrollados recientemente han integrado diversas señales para mejorar la sensibilidad en la identificación de *Indels*, aún presentan grandes limitaciones en la detección de estas variaciones de un rango de tamaño mayor, y la presencia de repeticiones cortan en tándem (STR) complica el mapeo dado que introducen ambigüedades dentro de una posición real en un *read* (Li, 2012). En base a ello, diversos estudios sugieren que para aumentar la precisión en la detección de *Indels* se utilice un conjunto de algoritmos (con ambos enfoques) con el fin de validar los resultados obtenidos y seleccionar aquellos programas con mejor desempeño (Hasan *y col.*, 2015; Xue *y col.*, 2016). De este modo se disminuyen las identificaciones erróneas (falsos positivos y negativos) en la detección de variantes de tipo *Indel* en secuencias de NGS, lo que permite mejorar la interpretación de la información genética de las muestras en estudio.

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*, 2013), 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).

En nuestro laboratorio se ha realizado secuenciación masiva de nueva generación a muestras de pacientes con adenocarcinoma gástrico, en primera instancia ya se han identificado variantes somáticas de tipo SNP para el gen MICA. Sin embargo, se hace necesario la utilización de herramientas bioinformáticas que den cuenta de la presencia de *Indels* en su secuencia, debido a la importancia que tendrían estas mutaciones en la funcionalidad del ligando. Por ello, el objetivo del presente proyecto es la utilización de diversos algoritmos de alineamiento local para la identificación de *Indels* en el gen *MICA*, comparando y evaluando las variantes identificadas para cada análisis. Tales *Indels* podrían estar dando cuenta de mecanismos de evasión inmune propios del desarrollo tumoral, mediante la interacción alterada de NKG2D y su ligando MICA en los pacientes en estudio. Por tal razón, 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.

1.8 Hipótesis:

"La utilización de diversos algoritmos de alineamiento local mejoran la precisión en la detección de *Indels,* implicados en la evasión inmune en cáncer gástrico".

1.9 Objetivo General:

Generar un protocolo para evaluar *Indels* detectados mediante secuenciación dirigida de alta profundidad (NGS) en muestras de pacientes con adenocarcinoma gástrico, que permitan el desarrollo de un nuevo blanco terapéutico.

1.10 Objetivos Específicos:

1. Identificar *Indels* detectados por el flujo de trabajo Amplicon TruSeq (Illumina) en gen MICA, en datos provenientes de muestras de pacientes con adenocarcinoma gástrico.
2. Detectar *Indels* utilizando diferentes algoritmos para alineamiento local en datos de NGS.
3. Análisis *in silico* de *Indels* identificados y su efecto en el ligando MICA.

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	18
	> 5 y < 10	21
	< o = a 10 cm	11
Clasidicación TNM	T1	6
	T2	4
	T3	12
	T4	28
H. pylori	+	22
	-	28

2.2. Secuenciación de muestras

Durante la gastrectomía se obtuvieron muestras frescas de tejido tumoral de aproximadamente 3 mm^3, 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.

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. 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).

Se realizó secuenciación dirigida de los exones 2, 3, 4 y 5 del gen *MICA* (Figura 1), se utilizó un kit TruSeq Custom Amplicon en el equipo MiSeq (Illumina Inc, EE.UU), ubicado en el Centro de Investigacion y Tratamiento del Cancer, Facultad de Medicina, Universidad de Chile. Para ello las regiones objetivo se cubrieron por 8 amplicones, los que se ingresaron mediante coordenadas cromosómicas en el programa 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.

2.3. Análisis Bioinformático

2.3.1. Identificación de *Indels* con Somatic Variant Caller (SVC)

Para el análisis de datos, los archivos de secuencia se subieron al servidor BaseSpace provisto por Illumina (https://basespace.illumina.com/home/index), el cual permite el almacenamiento y posterior 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 2), de acuerdo al flujo de trabajo integrado Amplicon TruSeq (Illumina Inc, EE.UU).

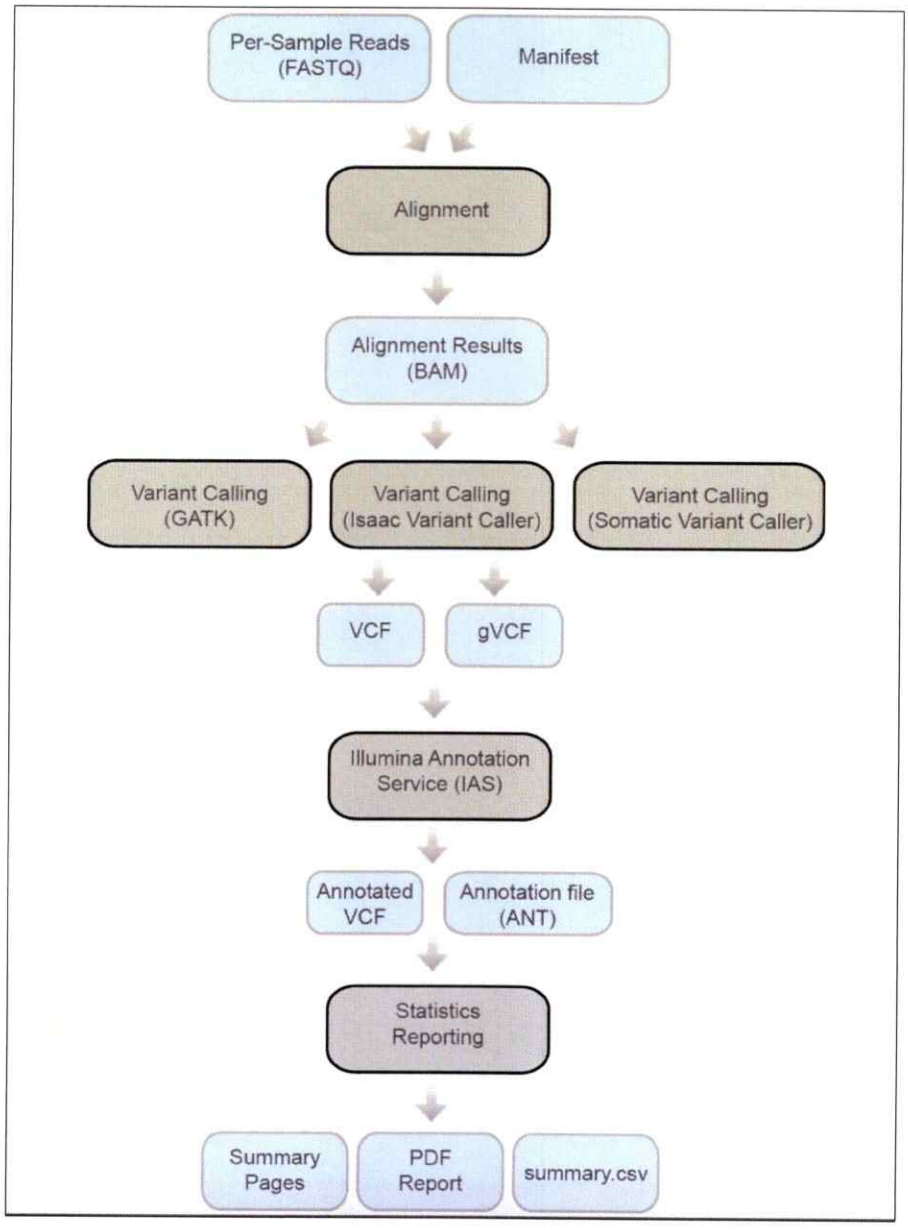


Figura 2. Herramientas informáticas utilizadas en el análisis de los datos de Illumina. Los archivos de secuencias (Fastq) generados durante la secuenciacion fueron alineados al genoma de referencia (Hg19). Para el llamado de variantes se utilizó el algoritmo Somatic Variant Caller (SVC) y las variantes 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.

La identificación de las variaciones (llamado de variante) presentes en el gen MICA para cada paciente se realizó utilizando el algoritmo Somatic Variant Caller (Figura 2), recomendado para el análisis de datos generados con el kit utilizado en este proyecto. 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).

Filtrado de las variantes detectadas

Si bien los archivo 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 2 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).

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 y referencia dbSNP, entre otros (Figura 3).

Tabla 2: Lista de filtros utilizados en la detección de variantes por SVC.

Tipo de Filtro	Efecto
"LowCoverage" 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.
"VeryLowQual" o filtro de muy baja calidad.	Elimina aquellas variaciones con una puntuación de calidad de menos de 30, que suelen ser artefactos
"LowQual" o filtro de baja calidad	Elimina aquellas variaciones con puntuaciones de calidad entre 30 y 50, que pueden ser artefactos.
"LowQD" 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.
"StrandBias".	Variaciones que sólo aparecen en las lecturas de la misma dirección son habitualmente artefactos, por lo que se filtran.

Sample	Gene	Variant	Chr	Coordinate	Type	Genotype	Exonic	Filters	Alt Variant Freq	Read Depth	Alt Read Depth	Allelic Depths	Num Transcripts	Transcript	Consequence
CG45	MICA	TG>TG/T	6	31380160	deletion	het	yes	PASS	2.28	3912	89	3823,89	1	NM_001177519.1	frameshift_variant, feature_tr
CG46	MICA	TG>TG/T	6	31380160	deletion	het	yes	PASS	96.95	4233	4104	1,294,104	1	NM_001177519.1	frameshift_variant, feature_tr
CG50	MICA	TG>TG/T	6	31380160	deletion	het	yes	PASS	2.7	4000	108	3,892,108	1	NM_001177519.1	frameshift_variant, feature_tr
CG58	MICA	TG>TG/T	6	31380160	deletion	het	yes	PASS	2.88	1907	55	1852,55	1	NM_001177519.1	frameshift_variant, feature_tr
CG01	MICA	G>G/GCTGCTGCT	6	31380161	insertion	het	yes	PASS	5.13	8933	458	8,475,458	1	NM_001177519.1	frameshift_variant, feature_el
CG01	MICA	G>G/GCTGCTGCTGCT	6	31380161	insertion	het	yes	PASS	34.97	8933	3124	58,093,124	1	NM_001177519.1	frameshift_variant, feature_el
CG02	MICA	G>G/GCT	6	31380161	insertion	het	yes	PASS	57.42	4220	2423	17,972,423	1	NM_001177519.1	frameshift_variant, feature_el
CG02	MICA	G>G/GCTGCTGCT	6	31380161	insertion	het	yes	PASS	4.6	4220	194	4,026,194	1	NM_001177519.1	frameshift_variant, feature_el
CG02	MICA	G>G/GCTGCTGCTGCT	6	31380161	insertion	het	yes	PASS	34.72	4220	1465	27,551,465	1	NM_001177519.1	frameshift_variant, feature_el
CG06	MICA	G>G/GCT	6	31380161	insertion	het	yes	PASS	46.91	7945	3727	42,183,727	1	NM_001177519.1	frameshift_variant, feature_el
CG08	MICA	G>G/GCT	6	31380161	insertion	het	yes	PASS	1.61	6132	99	6033,99	1	NM_001177519.1	frameshift_variant, feature_el
CG08	MICA	G>G/GCTGCTGCT	6	31380161	insertion	het	yes	PASS	4.99	6132	306	5,826,306	1	NM_001177519.1	frameshift_variant, feature_el
CG08	MICA	G>G/GCTGCTGCTGCT	6	31380161	insertion	het	yes	PASS	39.35	6132	2413	37,192,413	1	NM_001177519.1	frameshift_variant, feature_el
CG13	MICA	G>G/GCT	6	31380161	insertion	het	yes	PASS	3.99	7593	303	7,290,303	1	NM_001177519.1	frameshift_variant, feature_el
CG13	MICA	G>G/GCTGCTGCT	6	31380161	insertion	het	yes	PASS	11.05	7593	839	6,754,839	1	NM_001177519.1	frameshift_variant, feature_el
CG13	MICA	G>G/GCTGCTGCTGCT	6	31380161	insertion	het	yes	PASS	79.23	7593	6016	15,776,016	1	NM_001177519.1	frameshift_variant, feature_el
CG14	MICA	G>G/GCT	6	31380161	insertion	het	yes	PASS	38.22	8240	3149	50,913,149	1	NM_001177519.1	frameshift_variant, feature_el
CG15	MICA	G>G/GCT	6	31380161	insertion	het	yes	PASS	3.93	2263	89	2174,89	1	NM_001177519.1	frameshift_variant, feature_el
CG15	MICA	G>G/GCTGCTGCT	6	31380161	insertion	het	yes	PASS	10.69	2263	242	2,021,242	1	NM_001177519.1	frameshift_variant, feature_el
CG15	MICA	G>G/GCTGCTGCTGCT	6	31380161	insertion	het	yes	PASS	79.85	2263	1807	4,561,807	1	NM_001177519.1	frameshift_variant, feature_el
CG16	MICA	G>G/GCT	6	31380161	insertion	het	yes	PASS	4.82	1038	50	988,50	1	NM_001177519.1	frameshift_variant, feature_el
CG16	MICA	G>G/GCTGCTGCT	6	31380161	insertion	het	yes	PASS	11.08	1038	115	923,115	1	NM_001177519.1	frameshift_variant, feature_el
CG16	MICA	G>G/GCTGCTGCTGCT	6	31380161	insertion	het	yes	PASS	76.69	1038	796	242,796	1	NM_001177519.1	frameshift_variant, feature_el
CG18	MICA	G>G/GCT	6	31380161	insertion	het	yes	PASS	48.45	5757	2789	29,682,789	1	NM_001177519.1	frameshift_variant, feature_el
CG19	MICA	G>G/GCT	6	31380161	insertion	het	yes	PASS	3.85	6730	259	6,471,259	1	NM_001177519.1	frameshift_variant, feature_el
CG19	MICA	G>G/GCTGCTGCT	6	31380161	insertion	het	yes	PASS	11.03	6730	742	5,988,742	1	NM_001177519.1	frameshift_variant, feature_el
CG19	MICA	G>G/GCTGCTGCTGCT	6	31380161	insertion	het	yes	PASS	78.96	6730	5314	14,165,314	1	NM_001177519.1	frameshift_variant, feature_el
CG21	MICA	G>G/GCT	6	31380161	insertion	het	yes	PASS	33.94	3745	1271	24,741,271	1	NM_001177519.1	frameshift_variant, feature_el
CG21	MICA	G>G/GCTGCTGCT	6	31380161	insertion	het	yes	PASS	8.01	3745	300	3,445,300	1	NM_001177519.1	frameshift_variant, feature_el
CG21	MICA	G>G/GCTGCTGCTGCT	6	31380161	insertion	het	yes	PASS	54.42	3745	2038	17,072,038	1	NM_001177519.1	frameshift_variant, feature_el
CG22	MICA	G>G/GCT	6	31380161	insertion	het	yes	PASS	2.13	6586	140	6,446,140	1	NM_001177519.1	frameshift_variant, feature_el
CG22	MICA	G>G/GCTGCTGCT	6	31380161	insertion	het	yes	PASS	5.66	6586	373	6,213,373	1	NM_001177519.1	frameshift_variant, feature_el
CG22	MICA	G>G/GCTGCTGCTGCT	6	31380161	insertion	het	yes	PASS	40.22	6586	2649	39,372,649	1	NM_001177519.1	frameshift_variant, feature_el
CG23	MICA	G>G/GCT	6	31380161	insertion	het	yes	PASS	98.78	1720	1699	211,699	1	NM_001177519.1	frameshift_variant, feature_el

Figura 3. Visualización de archivo VCF con *Indels* detectados en el gen *MICA* con el algoritmo SVC. En azul se destaca la columna que indica el nombre del gen estudiado, en este caso *MICA*. En amarillo se destacan algunas características de las variantes de tipo Indels identificadas; su respectiva coordenada cromosómica y su condición de PASS si pasó los filtros utilizados en el llamado de variante, descritos en la Tabla 2.

2.3.2 Identificación de *Indels* con diversos algoritmos de alineamiento local

Con el objetivo de detectar *Indels* en la secuencia del gen *MICA*, se utilizaron algoritmos ampliamente utilizados en la literatura para la detección de este tipo de variantes. Los programas seleccionados fueron: Pindel (Ye *y col.*, 2009) y SOAPindel (Li *y col.*, 2013) con un flujo de trabajo de *Breakpoint*, y Scalpel (Narzisi *y col.*, 2014), y VarScan2 (Koboldt *y col.*, 2012) de tipo *Soft-clip* (Tabla 3). Para el analisis se utilizó una *workstation* Lenovo ThinkStation P700, con 64 GB de RAM, Procesador Intel Xeon de 32 núcleos, 2.4 GHz, CentOS 6.7 Corporativo.

Tabla 3: Detalle de algoritmos utilizados

Programa	Versión	Método de análisis	Tipo de variante que identifica
Pindel	0.2.0	Punto de quiebre (*Breakpoint*)	Inserción, deleción, *Indels* complejo y variantes estructurales
Scalpel	0.5.3	Ensamblaje local	Inserción, Deleción
SOAPindel	2.1	Punto de quiebre (*Breakpoint*)	Inserción, Deleción e Inversiones
VarScan2	2.3.9	Ensamblaje local	Inserción, Deleción

Previo a la identificación de *Indels* en la secuencia del gen *MICA*, se utilizó un ADN comercial HD701 (Horizon Dx, Cambridge, UK) con el fin de determinar el correcto flujo de trabajo de los algoritmos seleccionados.

Este ADN contiene en su secuencia una deleción de 15 nucleótidos en el exón 19 del gen EGFR (Figura 4), por lo que se utilizó como una muestra de referencia para probar la precisión en el flujo de trabajo NGS de cada algoritmo utilizado

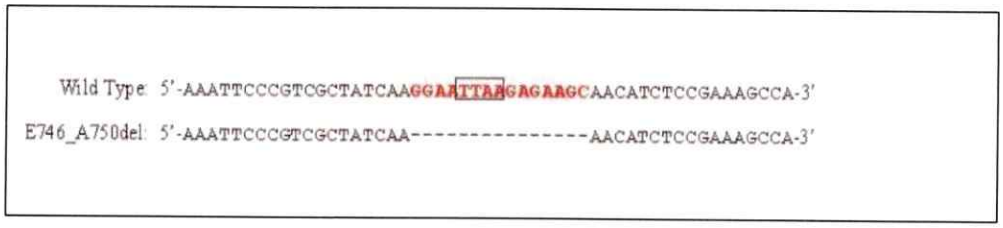


Figura 4. Deleción en el exón 19 del gen *EGFR* en la secuencia del ADN Horizon. En la imagen se muestra la secuencia wild type alineada a la secuencia que presenta la mutación E746_A750, correspondiente a una deleción. En rojo se destaca la secuencia de 15 nucleótidos que no están presentes en el exón 19 del ADN control (indicado con guiones).

Una vez identificado el *Indel* en el ADN control, se procedió a la identificación de *Indels* por cada algoritmo por separado, en las 50 muestras en estudio.

Comparación de *Indels* detectados por los algoritmos seleccionados.

Para comparar los resultados obtenidos por el grupo de algoritmos utilizado, se generó un diagrama de Venn, mediante una herramienta web provista por la Universidad de Ghent (http://bioinformatics.psb.ugent.be/webtools/Venn/). En el análisis se consideraron las coordenadas cromosómicas de los *Indels* detectados en el gen *MICA* por cada algoritmo en las muestras de CG en estudio.

A continuación se realizó un diagrama de Ventana deslizable (Bansal, 2012) con el programa RStudio versión 3.2.5. Donde a partir de una región de 10 nucleótidos denominanda "ventana" que se desplaza a través de la secuencia del gen *MICA*, se puede determinar cuantos *Indels* son detectados en la ventana por el grupo de algoritmos utilizados (Figura 5).

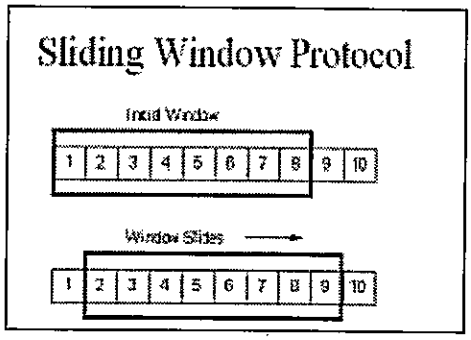


Figura 5. Protocolo de ventana deslizable. En la imagen se muestra una secuencia de largo 10 donde se seleccionan 8 posicionas que correspondiente a la región denominada "ventana". En la siguiente de abajo se muestra el desplazamiento de la ventana, avanzando una posición a la vez, a través de la secuencias con dirección hacia la derecha (indicado por la flecha).

Verificación mediante secuenciación de tipo Sanger de *Indels* detectados por NGS

Luego del análisis de NGS, los *Indels* detectados en MICA en pacientes con CG fueron validados mediante secuenciación de Sanger con el fin de realizar un análisis ortogonal de los resultados obtenidos, utilizando una técnica distinta de la utilizada inicialmente.

Se diseñó un par de partidores con el programa Amplifix (versión 1.5.4) que amplifican la región adyacente al los *Indels* detectados en la coordenada chr6:31380161.

Los partidores diseñados fueron MICAamp-F: 5'-TGCTGGTGCTTCAGAGTCAT -3' y MICAamp-R: 5'- AAGCCTTGTCACCAACATGC-3', los que amplificaron un fragmento de 231 pb que incluye al exón 5 de MICA codificante para la región TM, donde se han identificados del triplete GCT.

La reacción de PCR se realizó en un volumen de reacción de 50 ul y se utilizó la enzima polimerasa Platinum Taq DNA Polymerase High Fidelity (Invitrogen, USA). El ciclo de PCR incluyó una denaturación inicial a 96°C por 2 min, seguido de 10 ciclos de 96°C

durante 30 s, 68°C durante 50 s, 96°C durante 30 s, 62°C por 1 min y 72°C durante 2 min. Finalmente se realizó una extensión adicional de 72°C durante 10 min y las muestras se mantuvieron a 4°C hasta los próximos análisis.

Los productos de PCR se purificaron con el kit Wizard SV Gel and PCR Clean-Up System (Promega, USA), y se visualizaron en un gel de agarosa al 1%. Aproximadamente 50 ng fueron secuenciados mediante la metodología de Sanger en el equipo ABI PRISM 3500 xl (Applied Biosystems, USA), ubicado en el Centro de Secuenciación Automática de ADN, de la PUC. Los electroferogramas de secuencia se analizaron de manera manual con el programa DNASTAR (Version 14.1.0.115).

Asignación de variantes MICA-STR

La caracterización de *MICA* de acuerdo a la secuencia de su región TM se basa en el número de unidades repetidas GCT (triplete que codifica para el aminoácido Alanina) presentes en los productos de PCR que amplifican al exón 5 del gen. Las variantes podrán tener 4, 5, 6 o 9 repeticiones de la secuencia GCT (STR)(identificándolos como variantes MICA-A4, MICA-A5, MICA-A6 o MICA-A9, respectivamente). Adicionalmente, variantes con 5 repeticiones de GCT podrían presentar una inserción adicional de 'G' y son referidos como MICA-A5.1

Comparación entre algoritmos

De acuerdo a los resultados obtenidos mediante Sanger se procedió a evaluar la deteccion de variantes de tipo *Indels* con los algoritmos utilizados. Para ello se seleccionaron los parámetros Rellamado (que se define como la probabilidad que un

evento real sea identificado) y la Precisión (probabilidad que evento identificado sea real) descritos a continuación:

$$\text{Rellamado} = \frac{TP}{TP + FN}$$

$$\text{Precisión} = \frac{TP}{TP + FP}$$

Hasan y *col.*, 2015

Donde TP (verdadero positivo) corresponde a I*ndels* identificados por Sanger y por NGS, FP (falso positivo) corresponde a *Indels* detectados por NGS y no por Sanger, y FN (falso negativo) corresponde a *Indels* que no son detectados por NGS.

2.4. Predicción del efecto de *Indels* en la proteína MICA

Base de datos COSMIC y BrowserBeta

Con el objetivo de determinar si los *Indels* detectados mediante NGS ya habían sido descritos para la proteína MICA, se utilizó la base de datos COSMIC (http://cancer.sanger.ac.uk/cosmic) que recogen mutaciones descritas en cáncer alrededor del mundo y la base de datos BrowserBeta que presenta información acerca de secuencias exónicas generadas de diversos proyecto genómicos (http://exac.broadinstitute.org/)

Algoritmo SIFTIndel

Con el fin de determinar el efecto de variaciones en la proteína MICA, se utilizó el algoritmo disponible desde la web SIFTIndel v1.03 (Sorting Tolerant From Intolerant) (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.

2.5. 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 6). Este programa permite diferenciar visualmente detalles de la variantes analizadas (como cobertura de la región y alineamiento de lectura) y facilita la comparación entre muestras.

Figura 6. Ejemplo de visualización de los datos con el programa IGV. Se muestra el área cromosómica correspondiente y en la parte superior muestra un histograma con la profundidad de cobertura de las lecturas para cada posición.

RESULTADOS

3.1 Detección de *Indels* en el gen de *MICA* utilizando el flujo de trabajo Amplicon TruSeq (Illumina).

Mediante un estudio de secuenciación masiva del gen *MICA* en muestras de CG con la plataforma MiSeq de Illumina, se detectaron variaciones de un nucleótido (SNV) en su secuencia. El análisis bioinformático de las secuencias fue realizado mediante el flujo de trabajo Amplicon TruSeq, que utiliza el algoritmo Somatic Variant Caller (SVC) (Illumina) (Gárate-Calderón y col., en preparación). En Gárate-Calderón, 2016 se describió una nueva variante, P294A, la que, en relación a los datos entregados por la base de datos SNPeffect, presentó una reducción en la estabilidad de la proteína MICA por su localización, en una región de posibles cortes proteolíticos mediados por metaloproteasas. Tales 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.

En esta tesis, y dado que la proteína MICA es altamente polimórfica, se analizó la presencia de *Indels* en su secuencia, los que de acuerdo a diversos estudios podrían estar

afectando la conformación y función de este ligando en células NK (Martinez-Chamorro *y col.*, 2016).

La corrida de secuenciación de las 50 muestras en estudio incluyó otros genes relacionados a MICA, como lo son MICB y los genes ULBP 1-6, sin embargo los datos generados para tales genes no fueron analizados en el presente estudio. Del total de lecturas generadas (10.889.956 lecturas) el 95.5% pasó los filtros de castidad de Illumina, estos filtros evalúan la calidad de la imagen en el llamado de bases asignandole 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. Ante esto, 10.545.717 lecturas pasaron los filtros de castidad, pudiendo ser utilizadas para el posterior llamado de variante.

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 4.500x, dando cuenta que en general cada una de las bases de las regiones en estudio se leyeron más de 4.500 veces y que tal valor asegura que con los datos de secuencia obtenidos es posible realizar análisis informáticos para identificar variantes.

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. 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), valor que se considera óptimo para realizar el llamado de variante (Yost y col., 2012)..

A través del análisis de los datos de NGS del gen *MICA* utilizando el algoritmo SVC, se detectaron 8 variaciones de tipo *Indel* (Figura 7), 3 inserciones y 5 deleciones. Se observó además, una distribución heterogénea de las variaciones identificadas a lo largo del gen *MICA,* debido a que se observan variaciones tanto en exones como intrones del gen en estudio (Figura 8).

Gene	Variant	Chr	Coordinate	Type	Genotype	Exonic	Num Transcripts	Transcript	Consequence
MICA	A>A/AC	6	31378770	insertion	het	no	1	NM_001177519.1	intron_variant, feature_elongation
MICA	AT>AT/A	6	31379277	deletion	het	no	1	NM_001177519.1	intron_variant, feature_truncation
MICA	TG>TG/T	6	31379887	deletion	het	yes	1	NM_001177519.1	frameshift_variant, feature_truncation
MICA	TG>TG/T	6	31380001	deletion	het	yes	1	NM_001177519.1	splice_region_variant, feature_truncatio
MICA	C>C/CT	6	31380091	insertion	het	yes	1	NM_001177519.1	intron_variant, feature_elongation
MICA	TGCTG>TGCTG/T	6	31380157	deletion	het	yes	1	NM_001177519.1	frameshift_variant, feature_truncation
MICA	TG>TG/T	6	31380160	deletion	het	yes	1	NM_001177519.1	frameshift_variant, feature_truncation
MICA	G>G/GCTGCTGCT	6	31380161	insertion	het	yes	1	NM_001177519.1	frameshift_variant, feature_elongation

Figura 7. Detalle de los *Indels* detectados en el gen MICA con el algoritmo SVC. En amarillo se destacan algunas características de las variantes de tipo *Indels* identificadas; su respectiva coordenada cromosómica y su tipo si corresponde a una inserción o deleción. En rojo se destacan las variantes localizadas en regiones intrónicas en el gen *MICA*; en verde se destacan aquellas localizadas en regiones exónicas que generan un desplazamiento del marco de lectura.

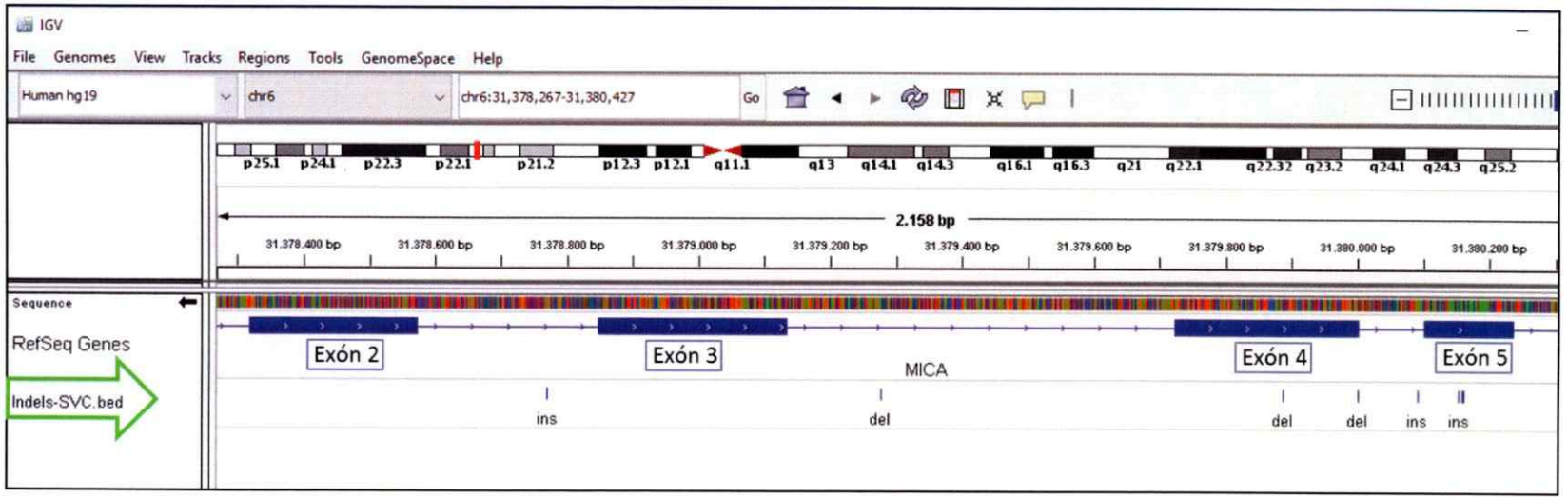


Figura 8. Distribución de los *Indels* detectados por el algoritmo SVC para el gen *MICA*, en muestras de pacientes con adenocarcinoma gástrico. El gen *MICA* se representa por bandas azules, los I*ndels* distribuidos a lo largo del gen se presentan en la fila señalada por la flecha verde. Las variaciones fueron visualizadas con el programa IGV versión 2.3.

3.2 Detección de *Indels* en el gen *MICA* mediante diversos algoritmos para alineamiento local.

Dado que el algoritmo SVC no presenta en su flujo de trabajo pasos de realineamiento de *Indel* o "limpieza de *Indels*" en el análisis de NGS, se consideró utilizar un grupo de Indel Callers (algoritmos llamadores de *Indels*), que realizan alineamiento local en el análisis de NGS. Los programas seleccionadas fueron Pindel, Scalpel, SOAPIndel y Varscan2.

3.2.1 Evaluación de *Indels* detectados en ADN Control

En primer lugar, se realizó la detección de *Indels* en el ADN control HorizonDX, por los 4 algoritmos seleccionado. Esta muestra control fue secuenciada previamente en el laboratorio y su particularidad radica en que es ampliamente utilizado como ADN de referencia para análisis de datos NGS, debido a que presenta una serie de variantes genéticas previamente descritas. Por tal razón, la utilización de este ADN control permite establecer una variante de tipo deleción a ser detectada mediante el análisis bioinformático de los algoritmos evaluados.

Luego del análisis del ADN control, se tiene que todos los algoritmos seleccionados detectaron la deleción de 15 nucleótidos en el gen EGFR descrita. En particular, el algoritmo Pindel genera un archivo intermedio en su flujo de trabajo, permitiendo visualizar el alineamiento de secuencias de la muestra analizada; en este caso, se observa la deleción de 15 nucleótidos del ADN control (Figura 9). Los resultados obtenidos validan la utilización de los programas seleccionados en la identificación de deleciones en las secuencias de NGS.

Figura 9. Identificación de deleción en muestra ADN Control, por el algoritmo Pindel. El algoritmo Pindel genera un archivo intermedio de alineamiento de secuencias que permite visualizar la deleción detectada. En el cuadro azul se destaca la presencia de la deleción localizada en el exón 19 del ADN control analizado, en el cuadro rojo de destacan las características de la deleción detectada, en él se indica que se han eliminado 15 nucleótidos de la secuencia del ADN control.

3.2.2 Detección de *Indels* en la secuencia del gen *MICA*

Luego de detectar el *Indel* descrito en el ADN Control con los diversos algoritmos seleccionados, se procedió a analizar los datos las secuencias de NGS para el gen *MICA* en las muestras en estudio. Como resultado del análisis bioinformático, todos los algoritmos detectaron *Indels* en la secuencia de *MICA* (Tabla 4).

Tabla 4: Número de *Indels* detectados en la secuencia de *MICA*, por los algoritmos seleccionados

Algoritmo	Pindel	SOAP	SVC	Scalpel	Varscan
N° Indels	311	77	8	6	6

Análisis de concordancia de *Indels* detectados mediante diagrama de Venn

Con el fin de identificar los *Indels* que fueron comunes para el grupo de programas utilizados, se analizó la concordancia de los 5 algoritmos mediante un diagrama de Venn, de modo de representar gráficamente las variantes que son comunes y aquellas que sean particulares entre los diversos programas.

La concordancia de los *Indels* identificados fueron 5 coordenadas cromosómicas comunes por los 5 algoritmos utilizados (Figura 10). Los *Indels* detectados por Scalpel y Varscan presentaron la mayor concordancia general, dado que considerando las 5 coordenadas comunes, cada programa identificó en total 6 variantes. En contraste, los *Indels* detectados por Pindel presentaron la mayor discordancia con los demás algoritmos, identificando 244 variantes que son particulares para este algoritmo.

En la Figura 9 se observa que los algoritmos SVC, Scalpel y Varscan presentaron resultados altamente concordantes entre sí, dado que considerando los 5 *Indels* comunes, estos programas no superan los 8 *Indels* detectados.

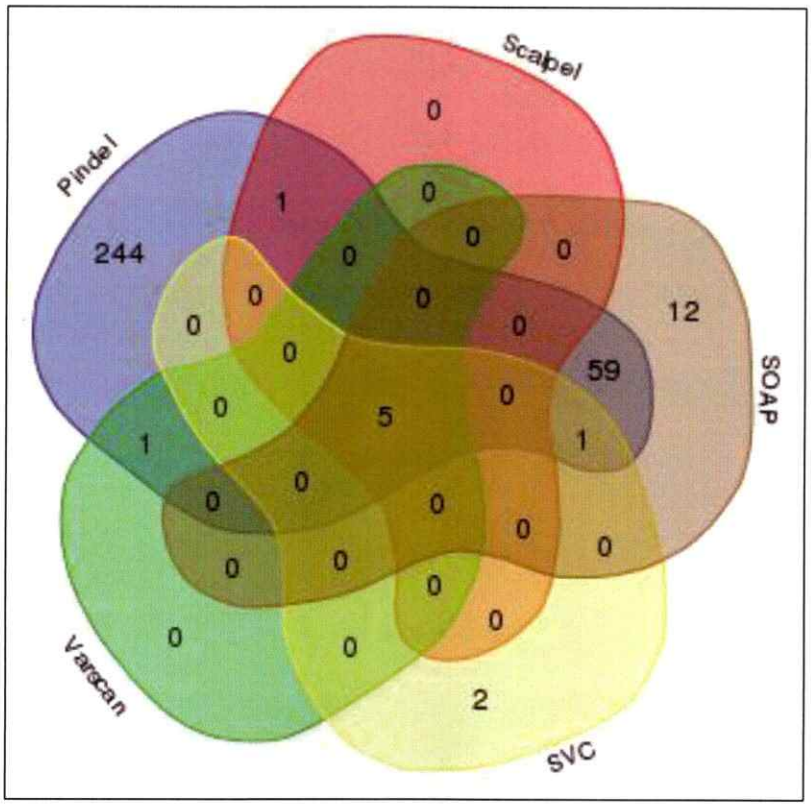


Figura 10. Concordancia de *Indels* detectados por Pindel, SOAPIndel, Scalpel, VarScan y SVC. En el centro se destacan 5 coordenadas cromosómicas que son comunes por los 5 programas utilizados. VarScan y Scalpel presentan alta concordancia en la identificación de *Indels*; considerando las 5 coordenadas comunes, cada programa identificó al menor 6 variantes. Pindel presenta la mayor discordancia con los demás algoritmos, identificando 244 variantes que son particulares para este algoritmo El diagrama fue creado con la herramienta web provista por la Universidad de Gent.

De los *Indels* adicionales a los 5 comunes, se tiene que sólo aquel detectado por Scalpel corresponde a un *Indel* descrito en la base de datos Browser Beta, chr6: 31378321 GC/C; por otro lado, el resto de los 3 y 1 *Indels* detectados por SVC y Varscan, respectivamente, no estan descritos en bases de datos.

Por su parte, los algoritmos Pindel y SOAP identificaron un alto número de variantes discordantes, detectando más de 300 y 70 *Indels* respectivamente.

Análisis de *Indels* mediante diagrama de Ventana Deslizable

Dada la disparidad en el número de *Indels* identificados por los diferentes algoritmos, se realizó una comparación de la *performance* de cada algoritmo mediante un diagrama de Ventana Deslizable, con el objetivo de determinar visualmente las regiones donde los programas estan identificando variantes a lo largo de la secuencia del gen en estudio.

El diagrama analiza mediante el desplazamiento de una Ventana de 10 nucleótidos a lo largo de la secuencia de *MICA*, la presencia de *Indels* detectados por cada algoritmo en esa determinada región (Figura 11). Los resultados del diagrama indican que los algoritmos Pindel y SOAPIndel están identificando un gran número de *Indels* en regiones cercanas a los *Indels* detectados por el resto de los algoritmos. En la Figura 11 se observa que Pindel y SOAPIndel, presentan un perfil que se visualiza como líneas más amplias a través de las ventanas fijadas en la secuencia de *MICA*. En tanto, para los algoritmos SVC, Varscan y Scalpel se visualizan líneas puntuales en regiones comunes con los 5 algoritmos utilizados. Estos resultados indican que a partir de las coordenadas cromosómicas que son comunes, los algoritmos menos restrictivos, Pindel y SOAPindel

(líneas azules y verdes a lo largo de la secuencia) están identificando una gran cantidad de *Indels* en las regiones cercanas, generando imprecisiones en la detección de estas variantes.

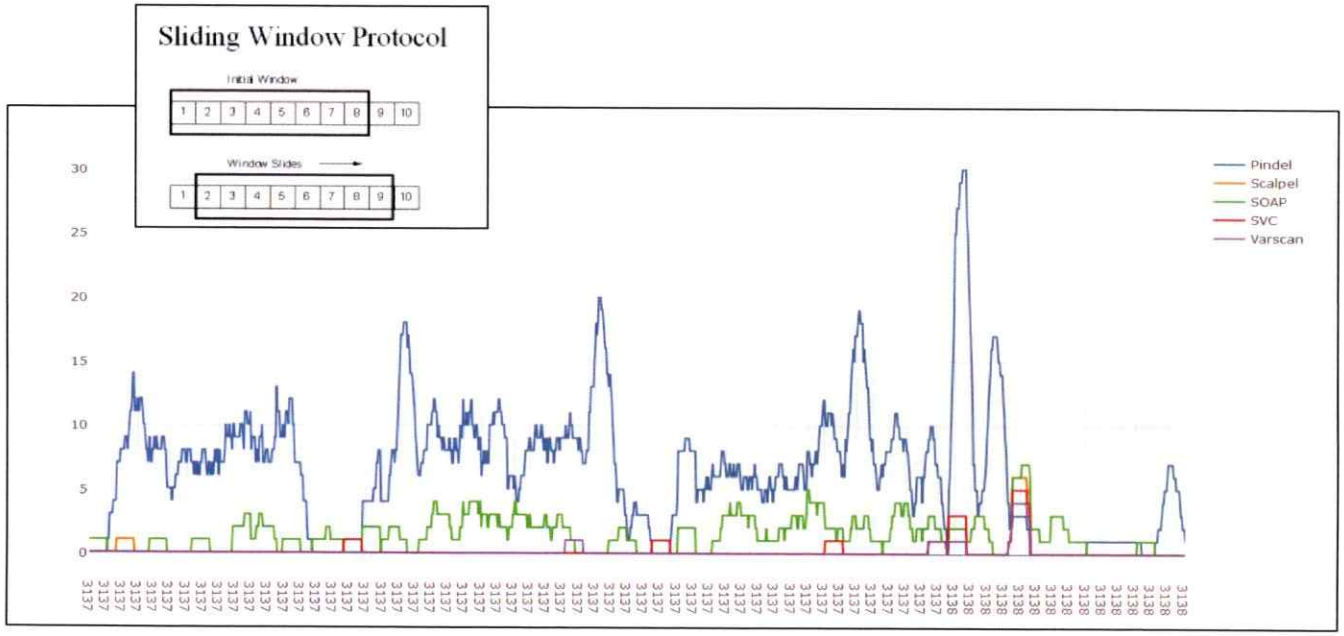


Figura 11. Diagrama de Ventana Deslizable para el llamado de Indels, a lo largo de la secuencia del gen MICA. El diagrama analiza mediante el desplazamineto de una ventana de largo definido la presencia de Indels en la secuencia de *MICA*, para cada uno de los programas utilizados en la llamada de variantes. En el caso de la secuencia de MICA se utilizó una Ventana de 10 nucleótidos. En el eje X, se presentan Ventana de 10 nucleótidos a lo largo de la secuencia del gen *MICA*, mientras que en eje Y se presenta el número de Indels detectados en esa determinada Ventana Diagrama realizado con el programa R-Studio versión 3.2.5.

Análisis de coordenadas cromosómicas comunes entre los algoritmos utilizados

Al realizar un análisis más detallado de las 5 coordenadas cromosómicas comunes para los 5 algoritmos utilizados, se tiene que dos de ellas están localizadas en la región intrónica del gen *MICA*, por lo que no estarían codificando para alterar la estructura de la proteína (Figura 12.A).

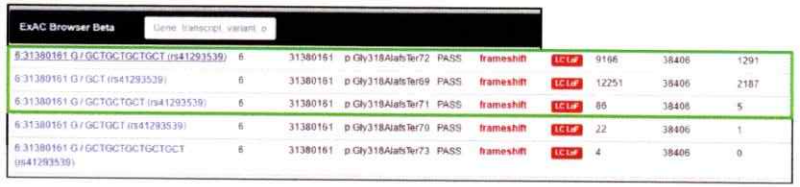
Adicionalmente como se muestra en la Figura 12 (Tabla A), para la coordenada chr6: 31380161 se detectan tres Inserciones con diferentes secuencias que inician en esta determinada posición (filas destacadas en celeste). Al observar en detalle la columna de las secuencias, se tiene que tales inserciones presentan en común un patrón de repeticiones del triplete GCT en su secuencia. Cabe señalar que el largo de las Inserciones analizadas no es divisibles por tres, por lo que estarían generando un desplazamiento en el marco de lectura, y en particular, la tabla muestra que generan un desplazamiento en la posición 318 de la proteína MICA.

Al realizar una búsqueda de estas variantes en diversas bases de datos genómicos, se tiene que ya han sido descritas tanto en Browser Beta (Figura 12.B), así como en la base de datos COSMIC (Figura 12.C). Precisando que tales variantes han sido reportadas en diversos tipos de cáncer, dado que alteran la estructura de la proteína MICA. En base a ello, se procedió a realizar un análisis detallado de la secuencia de estas variantes y de su efecto en la conformación de la proteína MICA.

A

Cromosoma	Coordenada	Gen	Ref	Alt	Tipo	Largo	Efecto	Cambio aminoácido	Efecto en proteína
chr6	31380001	MICA	TG	T	DEL	-1	splice_donor_variant&intron_variant	.	HIGH
chr6	31380091	MICA	C	CT	INS	1	splice_region_variant&intron_variant	.	LOW
chr6	31380157	MICA	TGCTG	T	DEL	-4	frameshift_variant	p.Gly318fs	HIGH
chr6	31380160	MICA	T	TG	INS	1	frameshift_variant	p.Gly318fs	HIGH
chr6	**31380161**	**MICA**	**G**	**GCTGCTGCTGCT**	**INS**	**11**	**frameshift_variant**	**p.Gly318fs**	**HIGH**
chr6	**31380161**	**MICA**	**G**	**GCTGCTGCT**	**INS**	**8**	**frameshift_variant**	**p.Gly318fs**	**HIGH**
chr6	**31380161**	**MICA**	**G**	**GCT**	**INS**	**2**	**frameshift_variant**	**p.Gly318fs**	**HIGH**

B



C

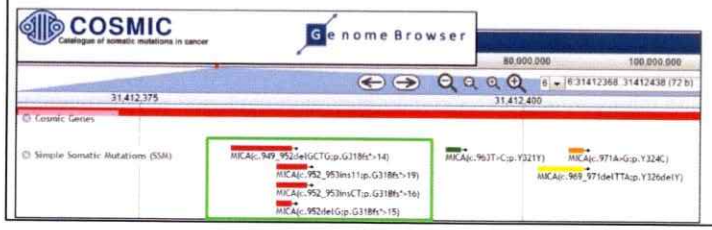


Figura 12. Detalle de *Indels* comunes para los 5 algoritmos utilizados, detectados en el gen MICA. Del grupo de 5 coordenadas cromosómicas detectados en el gen *MICA*, dos corresponden a regiones intrónica del gen y la coordenada chr6: 31380161 presenta tres secuencias diferentes para esa misma posición inicial, identificándose 3 Indels que presentan un perfil de repeticiones de la secuencia GCT. A) Tabla con detalle de los *Indels* detectados por los 5 algoritmos utilizados, en negrita se destaca la coordenada que presenta 3 Indels distintos para una misma coordenada inicial. B) Identificación de los Indels con repeticiones de GCT detectados en la base de datos Browser Beta. C) Identificación de los Indels con repeticiones de GCT en la base de datos COSMIC.

Al identificar el conjunto de inserciones localizadas en la coordenada chr6: 31380161, las que presentan repeticiones del triplete GCT en su secuencia, se tiene que los algoritmos utilizados estan llamando para una misma muestra inserciones con diferente largo de secuencia (Figura 13). Donde por ejemplo, para la muestra CG-53 el algoritmo Scalpel identifica inserciones con un largo de 2, 8 y 11 nucleótidos, mientras que para la misma muestra el algoritmo SOAPindel sólo identifica la insercion de largo 11.

Tal resultado podría estar indicando que para una misma muestra esten presente más de una secuencia con repetición de GCT en el ADN genomico analizado; o que por otro lado, que los algoritmos utilizados en el llamado de variante estan identificando secuencias parciales de una inserción más larga. Por tal razón, se procedió a realizar secuenciacion por Sanger de un grupo de muestras en estudio, con el fin de comprobar la presencia de tales inserciones en la secuencia del gen *MICA*.

Figura 13. Detección de las Inserciones localizadas en la coordenada chr6: 31380161 por los diversos algoritmos utilizados. En la imagen se delimita con un cuadro de color cada grupo de Inserciones localizadas en la coordenada chr6: 31380161, donde se detectan diferentes repeticiones de la secuencia GCT. Las repeticiones de GCT fueron detectadas por diferentes algoritmos, en este caso se visualizan los resultados de Scalpel (cuadro verde), SOAPindel (cuadro morado) y SVC (cuadro naranjo).

3.2.3 Validación de los *Indels* detectados en el gen *MICA*

Con el objetivo de validar la deteccion de *Indels* en secuencias de NGS, se evaluó la presencia de los *Indels* identificados mediante secuenciación de Sanger, verificando la presencia de estas variantes en los datos de secuencia de las muestras de cáncer gástrico analizadas.

A partir de los resultados del análisis bioinformático de NGS, se observó que las muestras de adenocarcinoma gástrico presentaron en su secuencia Inserciones con repeticiones de la secuencia GCT en la coordenada 31380161. Para la validación de estos resultados, se seleccionó un grupo de 5 muestras de CG con un perfil característico de la inserción con repetición de GCT (STR), que van desde 2 hasta 11 nucleótidos de largo y que se relacionan con variantes de MICA con codifican para diverso número de Alanina (A) en esta región. Las muestras seleccionadas fueron CG-06/14/29/33/40 (Figura 14).

Muestra	Programa	Cromosoma	Coordenada	Gen	Ref	Alt	Variante	Largo	Alelo MICA
CG06	Pindel	chr6	31380161	MICA	G	GCT	Inserción	2	A6
	Scalpel	chr6	31380161	MICA	G	GCT	Inserción	2	A6
	SOAP	chr6	31380161	MICA	G	GCT	Inserción	2	A6
	SVC	chr6	31380160	MICA	T	TG	Inserción	1	A5.1
	SVC	chr6	31380161	MICA	G	GCT	Inserción	2	A6
	Varscan	chr6	31380161	MICA	G	GCT	Inserción	2	A6
CG14	Programa	Cromosoma	Coordenada	Gen	Ref	Alt	Variante	Largo	Alelo MICA
	Scalpel	chr6	31380161	MICA	G	GCT	Inserción	2	A6
	SOAP	chr6	31380161	MICA	G	GCT	Inserción	2	A6
	SVC	chr6	31380160	MICA	T	TG	Inserción	1	A5.1
	SVC	chr6	31380161	MICA	G	GCT	Inserción	2	A6
	Varscan	chr6	31380161	MICA	G	GCT	Inserción	2	A6
CG29	Programa	Cromosoma	Coordenada	Gen	Ref	Alt	Variante	Largo	Alelo MICA
	Scalpel	chr6	31380160	MICA	TG	T	Inserción	1	A5.1
	Scalpel	chr6	31380161	MICA	G	GCTGCTGCT	Inserción	8	A8
	Scalpel	chr6	31380161	MICA	G	GCTGCTGCTGCT	Inserción	11	A9
	SOAP	chr6	31380160	MICA	TG	TGG	Inserción	1	A5.1
	SVC	chr6	31380160	MICA	T	TG	Inserción	1	A5.1
	SVC	chr6	31380161	MICA	G	GCT	Inserción	2	A6
	SVC	chr6	31380161	MICA	G	GCTGCTGCT	Inserción	8	A8
	SVC	chr6	31380161	MICA	G	GCTGCTGCTGCT	Inserción	11	A9
	Varscan	chr6	31380160	MICA	T	TG	Inserción	1	A5.1
	Varscan	chr6	31380161	MICA	G	GCTGCTGCTGCT	Inserción	11	A9
CG33	Programa	Cromosoma	Coordenada	Gen	Ref	Alt	Variante	Largo	Alelo MICA
	Scalpel	chr6	31380161	MICA	G	GCT	Inserción	2	A6
	Scalpel	chr6	31380161	MICA	G	GCTGCTGCT	Inserción	8	A8
	Scalpel	chr6	31380161	MICA	G	GCTGCTGCTGCT	Inserción	11	A9
	SOAP	chr6	31380161	MICA	G	GCTGCTGCTGCT	Inserción	11	A9
	SVC	chr6	31380160	MICA	T	TG	Inserción	1	A5.1
	SVC	chr6	31380161	MICA	G	GCT	Inserción	2	A6
	SVC	chr6	31380161	MICA	G	GCTGCTGCT	Inserción	8	A8
	SVC	chr6	31380161	MICA	G	GCTGCTGCTGCT	Inserción	11	A9
	Varscan	chr6	31380161	MICA	G	GCTGCTGCTGCT	Inserción	11	A9
CG40	Programa	Cromosoma	Coordenada	Gen	Ref	Alt	Variante	Largo	Alelo MICA
	Scalpel	chr6	31380161	MICA	G	GCTGCTGCT	Inserción	8	A8
	Scalpel	chr6	31380161	MICA	G	GCTGCTGCTGCT	Inserción	11	A9
	SOAP	chr6	31380161	MICA	G	GCTGCTGCTGCT	Inserción	11	A9
	SVC	chr6	31380161	MICA	G	GCT	Inserción	2	A6
	SVC	chr6	31380161	MICA	G	GCTGCTGCT	Inserción	8	A8
	SVC	chr6	31380161	MICA	G	GCTGCTGCTGCT	Inserción	11	A9
	Varscan	chr6	31380161	MICA	G	GCTGCTGCTGCT	Inserción	11	A9

Figura 14. Grupo de muestras seleccionadas para ser secuenciadas mediante secuenciación de Sanger. Del total de muestra con adenocarcinoma gástrico que presentaron Inserciones en la coordenada chr6:31380161, se seleccionaron 5 para ser secuenciadas mediante secuenciación por Sanger. Muestras seleccionadas CG-06/14/29/33/40.

3.2.3.1 Secuenciacion de Sanger a muestras seleccionadas

Se realizó un PCR a las muestras seleccionadas, con partidores dirigidos a amplificar la región chr6: 31380161, como se describe en Materiales y métodos. El tamaño del amplicón obtenido fue de 250 pb aproximadamente, cercano a los 231 pb del fragmento esperado (Figura 15). Posteriormente, los productos de PCR fueron secuenciados mediante secuenciacion por Sanger.

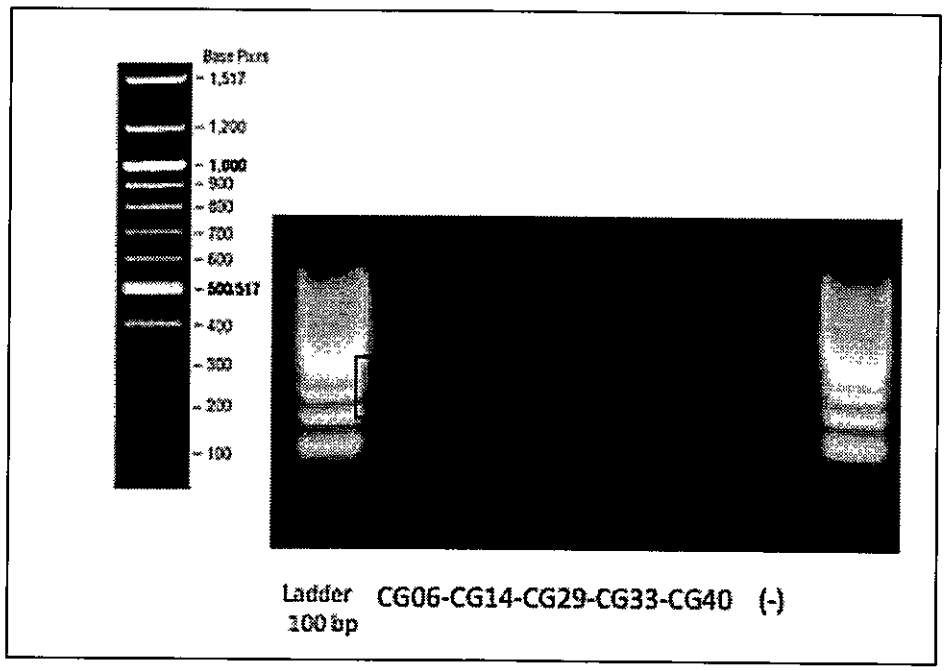
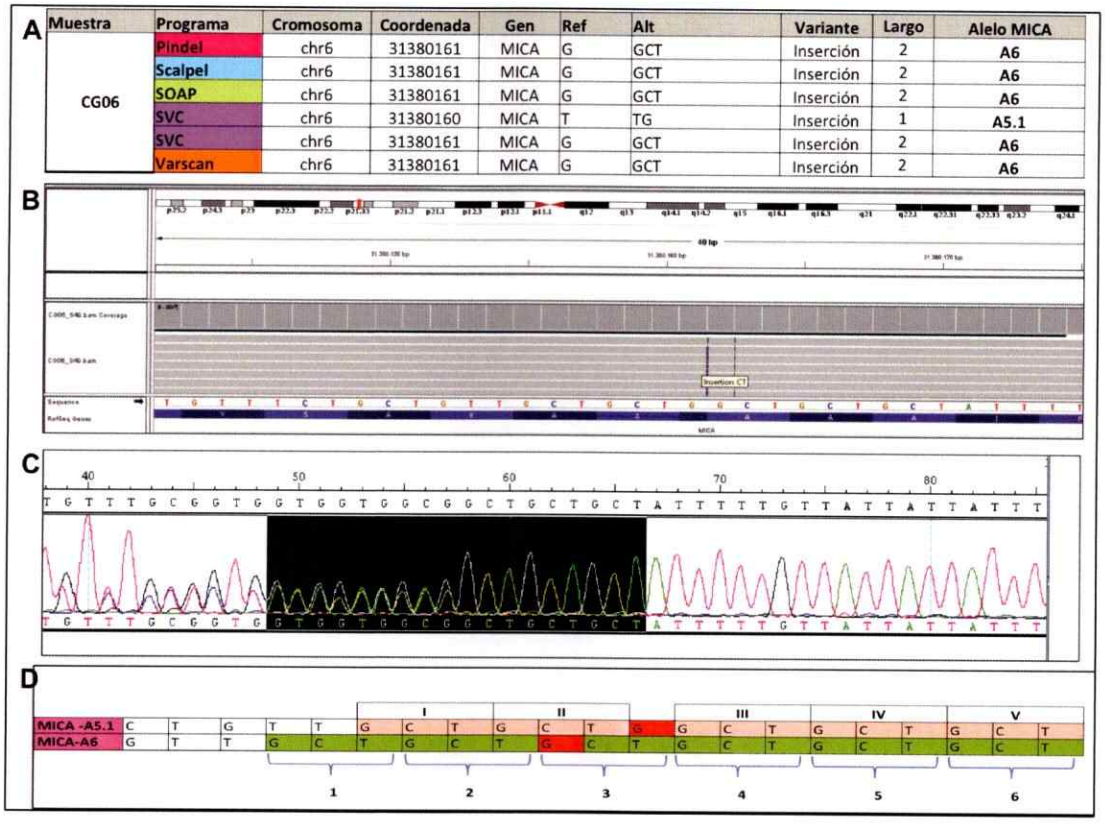


Figura 15. Visualización de productos de PCR que amplifican la región chr6: 31380161 en las muestras seleccionadas. Los productos de PCR fueron visualizados en un gel de agarosa al 1%, las bandas corresponden a la amplificacíon de la región cercana a la coordenada chr6: 31380161, con tamaño cercano a 250 pb aproximadamente.

A partir de los resultados de secuenciacion de Sanger se tiene que se identificaron las secuencias: 5.1A y 6A para las muestras CG-06 (Figura 16) y CG-14 (Figura 17); 5A y

9A para la muestra CG-29 (Figura 18); 6A y 9A para la muestra CG-33 (Figura 19) y 5.1A y 9A para la muestra CG-40 (Figura 20). Y se realizó un análisis comparativo entre los *Indels* detectado por Sanger y las variantes detectadas por NGS (Figura 21).



Muestra	Programa	Cromosoma	Coordenada	Gen	Ref	Alt	Variante	Largo	Alelo MICA
CG06	Pindel	chr6	31380161	MICA	G	GCT	Inserción	2	A6
	Scalpel	chr6	31380161	MICA	G	GCT	Inserción	2	A6
	SOAP	chr6	31380161	MICA	G	GCT	Inserción	2	A6
	SVC	chr6	31380160	MICA	T	TG	Inserción	1	A5.1
	SVC	chr6	31380161	MICA	G	GCT	Inserción	2	A6
	Varscan	chr6	31380161	MICA	G	GCT	Inserción	2	A6

Figura 16. Caracterización de la muestra CG-06 a partir del análisis de la secuenciación de Sanger. A) Detalle de los *Indels* identificados mediante análisis bioinformático en la muestra CG-06. B) Visualización de *Indels* en el programa IGV versión 2.3. C) Electroferograma con resultados de secuenciacion de Sanger. En negro se destaca el desplazamiento del marco de lectura para la secuencia del gen *MICA*. D) Análisis por separado de las secuencias del electroferograma. En rojo, se destaca la base que presenta una bifurcacion en los peaks de la secuencia del electroferograma; en naranjo, se muestra en la secuencia que se genera por el desplazamiento de lectura por la inserción de una base G (A5.1) y en verde se muestra la secuencia que presenta 6 repeticiones de GCT (A6).

A

	Programa	Cromosoma	Coordenada	Gen	Ref	Alt	Variante	Largo	Alelo MICA
CG14	Scalpel	chr6	31380161	MICA	G	GCT	Inserción	2	A6
	SOAP	chr6	31380161	MICA	G	GCT	Inserción	2	A6
	SVC	chr6	31380160	MICA	T	TG	Inserción	1	A5.1
	SVC	chr6	31380161	MICA	G	GCT	Inserción	2	A6
	Varscan	chr6	31380161	MICA	G	GCT	Inserción	2	A6

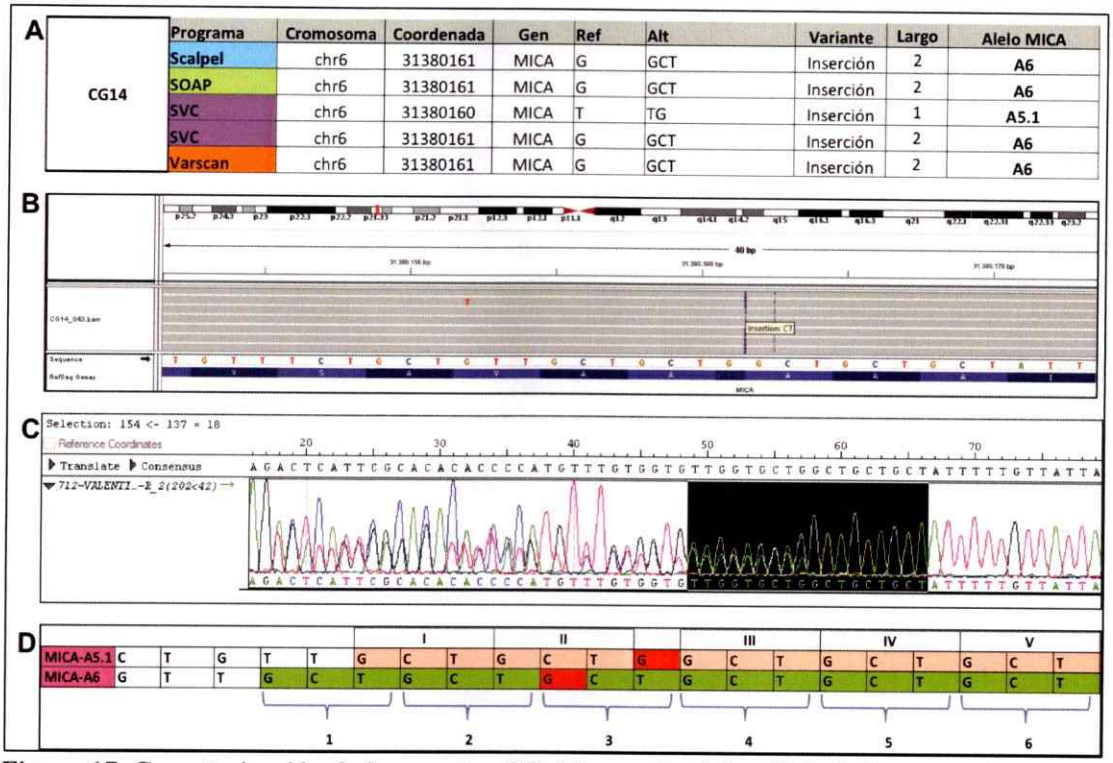


Figura 17. Caracterización de la muestra CG-14 a partir del análisis de la secuenciación de Sanger. A) Detalle de los *Indels* identificados mediante análisis bioinformático en la muestra CG-14. B) Visualización de *Indels* en el programa IGV versión 2.3. C) Electroferograma con resultados de secuenciacion de Sanger. En negro se destaca el desplazamiento del marco de lectura para la secuencia del gen *MICA*. D) Análisis por separado de las secuencias del electroferograma. En rojo, se destaca la base que presenta una bifurcacion en los peaks de la secuencia del electroferograma; en naranjo, se muestra en la secuencia que se genera por el desplazamiento de lectura por la inserción de una base G (A5.1) y en verde se muestra la secuencia que presenta 6 repeticiones de GCT (A6).

A

	Programa	Cromosoma	Coordenada	Gen	Ref	Alt	Variante	Largo	Alelo MICA
CG29	Scalpel	chr6	31380160	MICA	TG	T	Deleción	-1	A5.1
	Scalpel	chr6	31380161	MICA	G	GCTGCTGCT	Inserción	8	A8
	Scalpel	chr6	31380161	MICA	G	GCTGCTGCTGCT	Inserción	11	A9
	SOAP	chr6	31380160	MICA	TGG	TG	Deleción	-1	A5.1
	SVC	chr6	31380160	MICA	TG	T	Deleción	-1	A5.1
	SVC	chr6	31380161	MICA	G	GCT	Inserción	2	A6
	SVC	chr6	31380161	MICA	G	GCTGCTGCT	Inserción	8	A8
	SVC	chr6	31380161	MICA	G	GCTGCTGCTGCT	Inserción	11	A9
	Varscan	chr6	31380160	MICA	TG	T	Deleción	-1	A5.1
	Varscan	chr6	31380161	MICA	G	GCTGCTGCTGCT	Inserción	11	A9

Figura 18. Caracterización de la muestra CG-29 a partir del análisis de la secuenciación de Sanger. A) Detalle de los *Indels* identificados mediante análisis bioinformático en la muestra CG-29. B) Visualización de *Indels* en el programa IGV versión 2.3. C) Electroferograma con resultados de secuenciacion de Sanger. En negro se destaca el desplazamiento del marco de lectura para la secuencia del gen *MICA*. D) Análisis por separado de las secuencias del electroferograma. A partir de la bifurcacion en la secuencia; se observa en verde la secuencia con 9 repeticiones de GCT (A9); mientras que en la otra hebra, se muestra en naranjo la secuencia que se presenta como similar lo presentado en el genoma de referencia (A5).

A

	Programa	Cromosoma	Coordenada	Gen	Ref	Alt	Variante	Largo	Alelo MICA
CG33	Scalpel	chr6	31380161	MICA	G	GCT	Inserción	2	A6
	Scalpel	chr6	31380161	MICA	G	GCTGCTGCT	Inserción	8	A8
	Scalpel	chr6	31380161	MICA	G	GCTGCTGCTGCT	Inserción	11	A9
	SOAP	chr6	31380161	MICA	G	GCTGCTGCTGCT	Inserción	11	A9
	SVC	chr6	31380160	MICA	TG	T	Deleción	-1	A5.1
	SVC	chr6	31380161	MICA	G	GCT	Inserción	2	A6
	SVC	chr6	31380161	MICA	G	GCTGCTGCT	Inserción	8	A8
	SVC	chr6	31380161	MICA	G	GCTGCTGCTGCT	Inserción	11	A9
	Varscan	chr6	31380161	MICA	G	GCTGCTGCTGCT	Inserción	11	A9

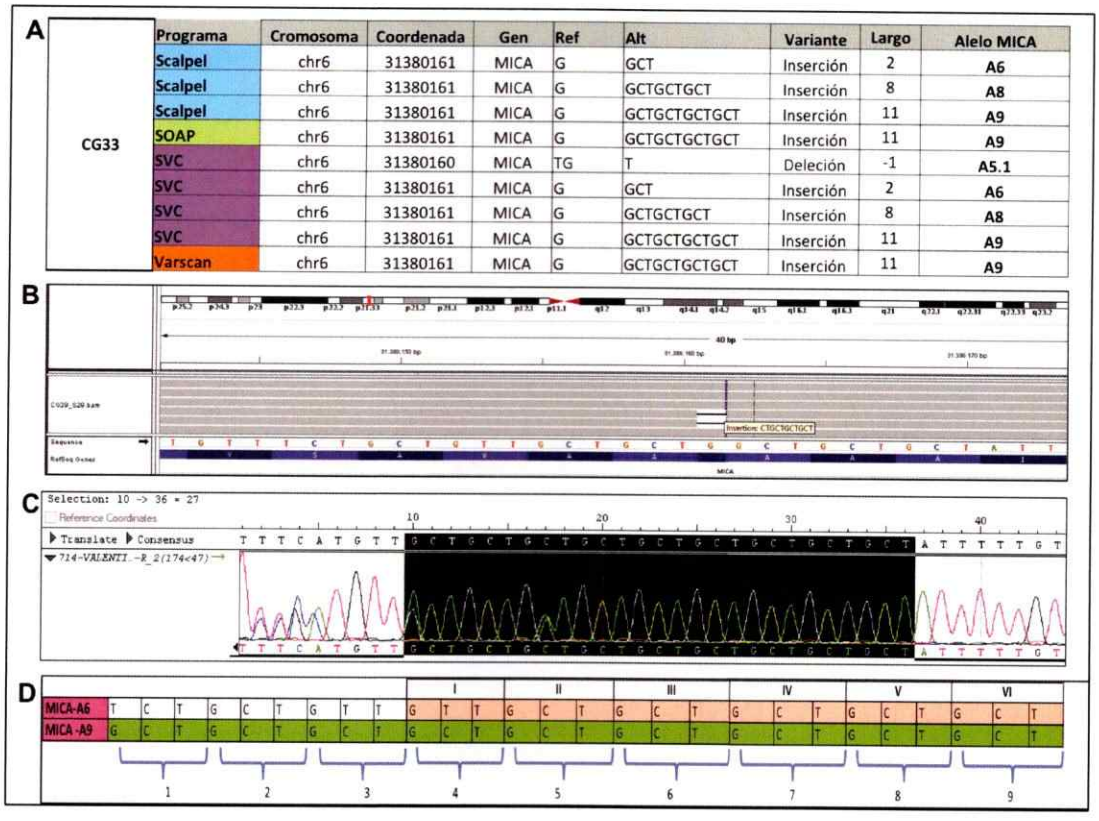


Figura 19. Caracterización de la muestra CG-33 a partir del análisis de la secuenciación de Sanger. A) Detalle de los *Indels* identificados mediante análisis bioinformático en la muestra CG-33. B) Visualización de *Indels* en el programa IGV versión 2.3. C) Electroferograma con resultados de secuenciacion de Sanger. En negro se destaca el desplazamiento del marco de lectura para la secuencia del gen *MICA*. D) Análisis por separado de las secuencias del electroferograma. A partir de la bifurcacion en la secuencia; se destaca en verde la secuencia con 9 repeticiones de GCT (A9) y en la otra hebra se presenta la secuencia que genera 6 repeticiones de GCT (A6).

A

	Programa	Cromosoma	Coordenada	Gen	Ref	Alt	Variante	Largo	Alelo MICA
CG40	Scalpel	chr6	31380161	MICA	G	GCTGCTGCT	Inserción	8	A8
	Scalpel	chr6	31380161	MICA	G	GCTGCTGCTGCT	Inserción	11	A9
	SOAP	chr6	31380161	MICA	G	GCTGCTGCTGCT	Inserción	11	A9
	SVC	chr6	31380161	MICA	G	GCT	Inserción	2	A6
	SVC	chr6	31380161	MICA	G	GCTGCTGCT	Inserción	8	A8
	SVC	chr6	31380161	MICA	G	GCTGCTGCTGCT	Inserción	11	A9
	Varscan	chr6	31380161	MICA	G	GCTGCTGCTGCT	Inserción	11	A9

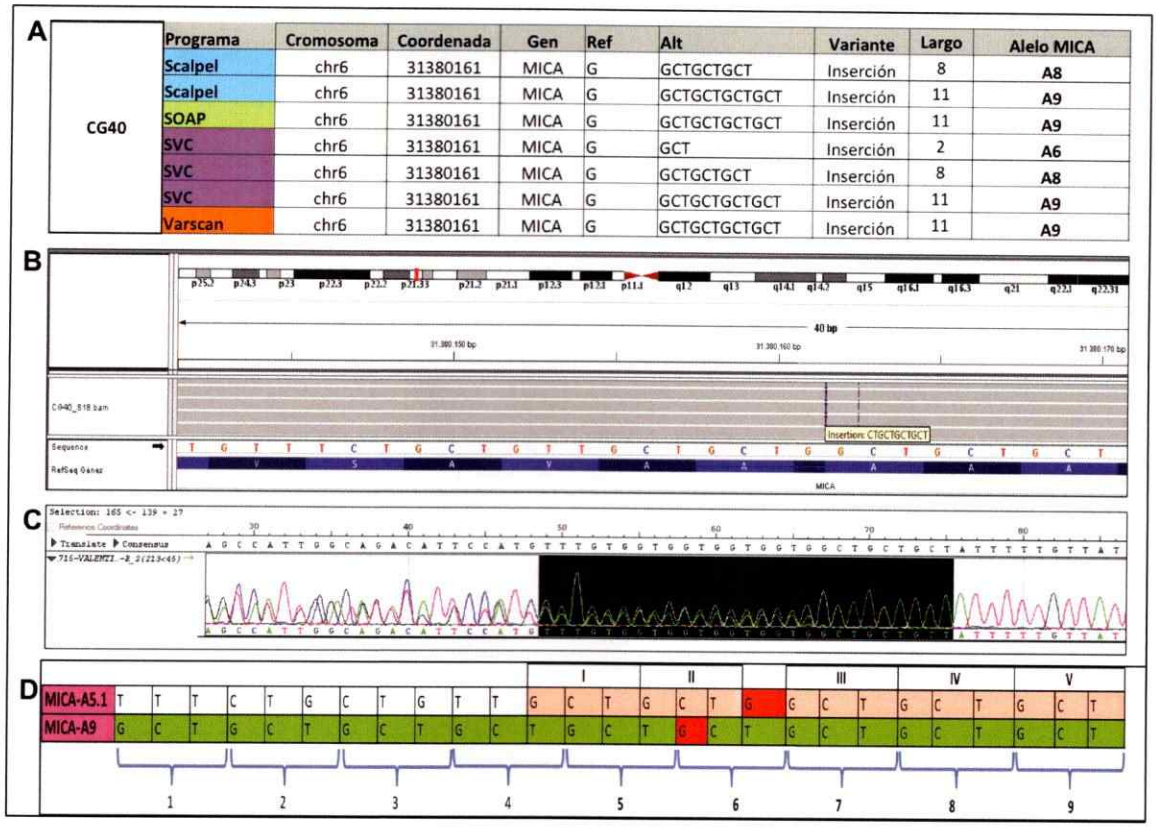


Figura 20. Caracterización de la muestra CG-40 a partir del análisis de la secuenciación de Sanger. A) Detalle de los *Indels* identificados mediante análisis bioinformático en la muestra CG-40. B) Visualización de *Indels* en el programa IGV versión 2.3. C) Electroferograma con resultados de secuenciacion de Sanger. En negro se destaca el desplazamiento del marco de lectura para la secuencia del gen *MICA*. D) Análisis por separado de las secuencias del electroferograma. En rojo, se destaca la base que presenta una bifurcacion en la secuencia; en verde se muestra la secuencia con 9 repeticiones de GCT (A9) y en la otra hebra en naranjo, se muestra en la secuencia que se genera por el desplazamiento de lectura por la inserción de una base G (A5.1).

GC-06	Resultado por Sanger: A5.1/ A6			
Algoritmo		TP	FP	FN
Pindel	A6	1	0	1
Scalpel	A6	1	0	1
SOAP	A6	1	0	1
SVC	A5.1/A6	2	0	0
Varscan	A6	1	0	1

GC-14	Resultado por Sanger: A5.1/ A6			
Algoritmo		TP	FP	FN
Pindel	.	0	0	2
Scalpel	A6	1	0	1
SOAP	A6	1	0	1
SVC	A5.1/A6	2	0	0
Varscan	A6	1	0	1

GC-29	Resultado por Sanger: A5/ A9			
Algoritmo		TP	FP	FN
Pindel	.	0	0	2
Scalpel	A5.1/A8/A9	1	2	0
SOAP	A5.1	0	1	1
SVC	A5.1/A6/A8/A9	1	3	0
Varscan	A5.1/A9	1	0	1

GC-33	Resultado por Sanger: A6/ A9			
Algoritmo		TP	FP	FN
Pindel	.	0	0	2
Scalpel	A6/A8/A9	2	1	0
SOAP	A9	1	0	1
SVC	A5.1/A6/A8/A9	2	2	0
Varscan	A9	1	0	1

GC-40	Resultado por Sanger: A5.1/ A9			
Algoritmo		TP	FP	FN
Pindel	.	0	0	2
Scalpel	A8/A9	1	1	1
SOAP	A9	1	0	1
SVC	A6/A8/A9	1	2	1
Varscan	A9	1	0	1

Figura 21. Comparación de la detección de *Indels* mediante resultados de NGS y secuenciación de Sanger para las muestras en estudio. Para cada una de las cinco muestras analizadas se muestra un cuadro comparatiivo, evidenciando en rosado lo detectado por secuenciacion por Sanger (considerado como verdadero) y bajo cada muestra se detalla lo obtenido por los algoritmos del análisis de NGS. TP, verdadero positivo (*Indel* detectado por Sanger y NGS); FP, falso positivo (*Indel* detectado por NGS y no por Sanger); FN, falso negativo (*Indel* detectado por Sanger y no por NGS); no se incluye TN, verdadero negativo, debido a que para la validación de la tecnología no.se consideraron pacientes sin *Indels* en el gen MICA.

Finalmente y con el objetivo de identificar los algoritmos con mejor desempeño en la identificación de *Indels* de tipo STR, se seleccionaron dos parámetros: Rellamado, que se define como la probabilidad que un evento real sea identificado y la Precisión, probabilidad que evento identificado sea real. Adicionalmente se consideró la Sensibilidad que se define como las variants verdaderas positivas identificadas en relación al total de variants identificadas por Sanger (Tabla 5).

Tabla 5: Detalle de parámetros analizados para los algoritmos

	Sensibilidad	Precisión	Rellamado
SVC	80%	0,53	0,88
Scalpel	60%	0,6	0,66
Varscan	50%	1	0,5
Pindel	10%	1	0,1
SOAPindel	40%	0,5	0,44

En base a estos resultados, los algoritmos Scalpel y SVC se presentan como los programas con mejores valores de sensibilidad, precisión y rellamado.

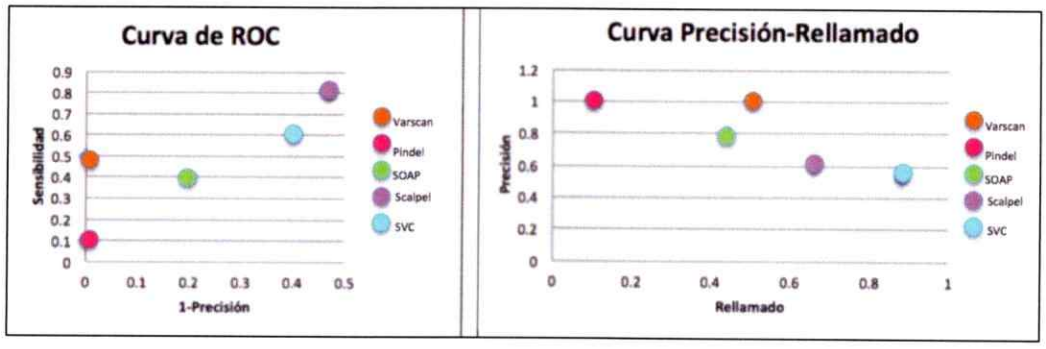


Figura 22. Comparación de parámetros en los algoritmos en estudio. Se muestra un análisis comparativo para el grupo de algoritmos analizados, mediante un gráfico de curva ROC y un gráfico de curva Rellamado-Precisión, evidenciando que los algoritmos SVC y Scalpel tiene alta sensibilidad y rellamado, sin embargo presentan una precisión media.

Sin embargo, al considerar el analisis comparativo mediante curvas de ROC y curva Rellamado-Precisión (Figura 22), se observa que los algoritmos con mejor desempeño (SVC y Scalpel) presentan una precisión media, por lo que se propone utilizar al menos tres de los programas utilizados con el fin de aumentar la precisión en la detección de *Indels* mediante NGS en la secuencia de *MICA* (Figura 23).

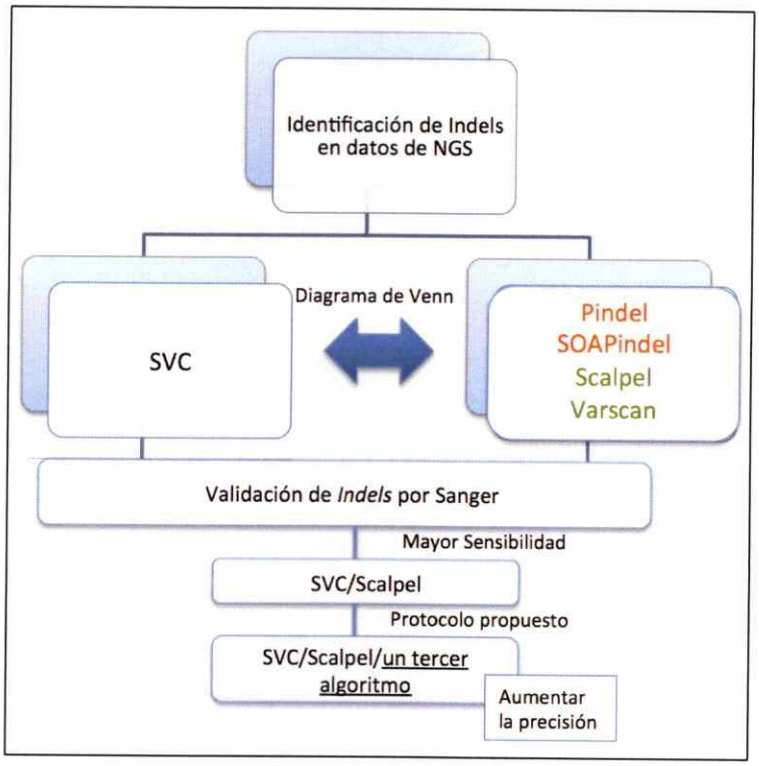
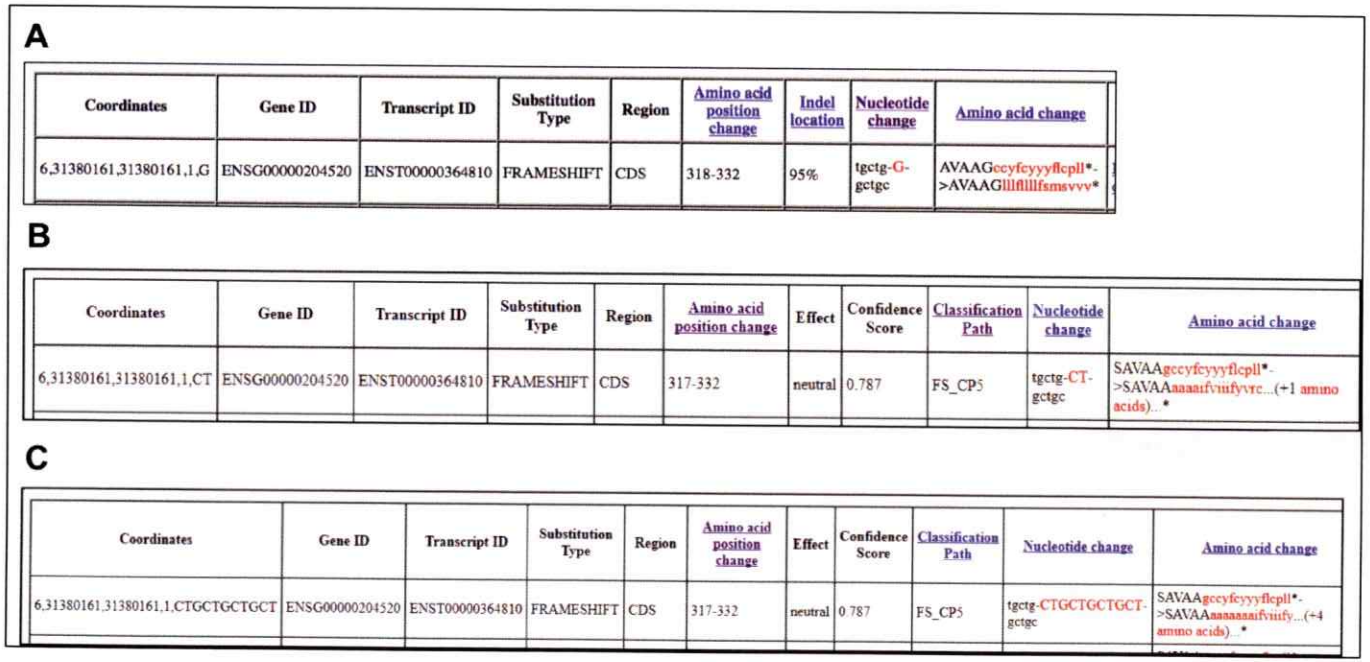


Figura 23. Diagrama del flujo de trabajo realizado por los cinco algoritmos seleccionados, en la identificación de *Indels*. Modelo de trabajo propuesto para la identificación de variantes de tipo *Indels* con STR, donde se propone utilizar los algoritmos SVC, Scalpel y un tercer programa de los utilizados, con el objetivo de aumentar la precisión en el llamado de variantes.

3.3 Análisis *in silico* del efecto de la Insercion chr6: 31380161 en el ligando MICA

Con el objetivo de determinar si los *Indels* identificados afectan la funcion del ligando MICA, se analizaron las variantes que codifican para 5.1, 6 y 9 repeticiones de Alaninas con el fin de predecir su efecto en la proteína.

En base a los resultados entregados por la base de datos SIFTIndel las variantes analizadas chr6:31380161,-/G (MICA-A5.1), chr6:31380161,G/CT (MICA-A6) chr6:31380161,G/CTGCTGCTGCT (MICA-A9), producen un cambio con efecto neutral en la proteína MICA. Si bien se detalla que las inserciones codifican para el aminoácido Alanina (1 y 4 aminoácido adicionales, respectivamente), este cambio no generaría una reducción en la estabilidad de la proteína (Figura 24). Adicionalmente la variante chr6:31380161,-/G (MICA-A5.1), generaría una proteína trunca en la posición 318 de la secuencia aminoacídica.



A

Coordinates	Gene ID	Transcript ID	Substitution Type	Region	Amino acid position change	Indel location	Nucleotide change	Amino acid change
6,31380161,31380161,1,G	ENSG00000204520	ENST00000364810	FRAMESHIFT	CDS	318-332	95%	tgctg-G-gctgc	AVAAGccyfcyyyflcpll*->AVAAGlllflllllfsmsvvv*

B

Coordinates	Gene ID	Transcript ID	Substitution Type	Region	Amino acid position change	Effect	Confidence Score	Classification Path	Nucleotide change	Amino acid change
6,31380161,31380161,1,CT	ENSG00000204520	ENST00000364810	FRAMESHIFT	CDS	317-332	neutral	0.787	FS_CP5	tgctg-CT-gctgc	SAVAAgccyfcyyyflcpll*->SAVAAaaaaifviiifyvrc...(+1 amino acids)...*

C

Coordinates	Gene ID	Transcript ID	Substitution Type	Region	Amino acid position change	Effect	Confidence Score	Classification Path	Nucleotide change	Amino acid change
6,31380161,31380161,1,CTGCTGCTGCT	ENSG00000204520	ENST00000364810	FRAMESHIFT	CDS	317-332	neutral	0.787	FS_CP5	tgctg-CTGCTGCTGCT-gctgc	SAVAAgccyfcyyyflcpll*->SAVAAaaaaaaaifviiify...(+4 amino acids)...*

Figura 24. Predicción del efecto de *Indels* detectados en la estructura de la proteína MICA. En base a la predicción realizada por la base de datos SIFIIndel, las variantes generan un efecto neutral en la estabilidad de la proteína. A) Detalle Inserción chr6:3180161,-/G (MICA-A5.1) de acuerdo a SIFT. B) Detalle Inserción chr6:3180161,G/CT (MICA-A6), de acuerdo a SIFT. C) Detalle Inserción chr6:3180161,G/CTGCTGCTGCT (MICA-A9) de acuerdo a SIFT.

DISCUSIÓN

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 Tγδ, 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. Esto, debido a 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 (Ribeiro *y col.*, 2016; Ghadially *y col.*, 2017).

Debido a lo anterior, en el presente estudio se evaluaron diversos algoritmos en la detección de variaciones genéticas de tipo *Indels* en el gen *MICA*, cuya proteína actúa como ligando de NKG2D. Para ello se analizaron secuencias obtenidas de pacientes con CG y se estudiaron los efectos que estos cambios podrían tener en la función de la proteína.

4.1 Detección de *Indels* en el gen *MICA*

MICA es un gen altamente polimórfico, a la fecha se conocen 105 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 variaciones genéticas de tipo *Indel* en regiones codificantes del gen *MICA,* se secuenció un grupo de 50 muestras tumorales provenientes de pacientes con adenocarcinoma gástrico. Sin embargo, la detección de *Indels* desde datos de NGS es un proceso complejo dado que para obtener los datos de secuencia de la región en estudio, los *reads* se alinean frente a un genoma de referencia y, de este modo, variantes de tipo *Indel* que conllevan la ganancia o pérdida de bases son especialmente proclives a ser mal alineadas contra el genoma de referencia produciendo una elevada tasa de falsos positivos durante el llamado de estas variantes (Breder *y col.*, 2017). Ante este escenario, diversos autores sugieren que para aumentar la precisión en la identificación de *Indels* es necesaria la validación de los resultados mediante la comparación de diversos algoritmos (Ghoneim *y col.*, 2014; Hasan *y col.*, 2015), e incluir en el flujo de trabajo bioinformático la detección de variantes desde datos de secuencias de ADN control de muestras comerciales (*gold standard*), con el fin de evaluar la *performance* de cada programa previo a su utilización en las muestras a ser analizadas (Hwang *y col,*. 2015).

En base a ello y con el fin de detectar este tipo de variantes en la secuencia del gen *MICA,* en el estudio se consideraron 5 algoritmos ampliamente utilizados en la literatura

para la detección de *Indels*. En primer lugar, y con el objeto de evaluar la sensibilidad en la detección de *Indels* de cada algoritmo, se utilizó un ADN control que presenta una deleción descrita para el gen *EGFR* (Figura 4); en base a la *performance* en la detección de la deleción desde la secuencia de este ADN, se validó la utilización de los programas SVC, VarScan2, SOAPIndel, Pindel y Scalpel en el estudio.

A continuación se trabajó con cada programa por separado, preparando inicialmente el análisis bioinformático para una muestra de CG y una vez determinado el flujo de trabajo para cada programa, se procedió a la identificación de *Indels* en la secuencia del gen *MICA* para las 50 muestras en estudio.

Como se muestra en la Figura 9, se visualizaron 5 coordenadas comunes entre los 5 programas utilizados; sin embargo, se hace evidente que hay programas más restrictivos que otros en la identificación de variantes de tipo *Indels* en las muestras analizadas. Por un lado tenemos a VarScan, SVC y Scalpel, que presentaron una alta concordancia entre sí y cada uno de ellos llama menos de 8 *Indels* en la secuencia del gen *MICA*. Y al analizar en detalle los *Indels* adicionales, se tiene que sólo aquel identificado por Scalpel ya ha sido descrito previamente en la base de datos Browser Beta, mientras que el resto de los Indels 3 y 1, detectados por SVC y Varscan respectivamente, no aparecen en bases de datos y su localización es cercana a SNPs ya descritos para el gen *MICA*; en base a ello, estos *Indels* podría estar dando luces de artefactos del alineamiento local realizado por estos algoritmos (Li, 2014), en regiones donde efectivamente se detectan variaciones de un nucleóido para el gen.

Por otra parte, se tiene que los programas Pindel y SOAPIndel identifican para la misma secuencia del gen *MICA* más de 300 y 70 *Indels*, respectivamente. Esta disparidad en los

resultados de *Indels* identificados por los diversos programas ya ha sido descrita previamente por Ghoneim *y col.*, 2014 donde se destaca que dada las diferentes estrategias de alineamiento local utilizados por los algoritmos, la identificación de *Indels* para un mismo set de datos presentará discrepancias entre los programas utilizados. En base a estas diferencias, en esta tesis se plantea como hipótesis una línea de trabajo en que se selecciona un conjunto de programas con diversas estrategias de alineamiento local y se evaluan los *Indels* detectados en busca de la mayor concordancia entre los programas. Es por esta razón que en el estudio se seleccionaron un conjunto de 5 programas con características específicas en la búsqueda de un amplio rango de *Indels* en la secuencia del gen *MICA* (Tabla 3).

Con el objetivo de estudiar la distribución del gran número de *Indels* discordantes detectados por Pindel y SOAPIndel, se analizaron los datos de secuencia en un gráfico de Ventana Deslizable (Figura 11). En la figura se observa que Pindel y SOAPIndel identifican *Indels* en regiones contiguas a las coordenadas comunes entre los 5 algoritmos. Este fenómeno se debería a que estos dos algoritmos además de detectar *Indels* pequeños y repeticiones cortas en tándem (STR), permiten identificar variaciones estructurales, deleciones largas (>100 pb) e *Indels* complejos (eventos de co-ocurrencias de Inserciones y Deleciones en una determinada región del genoma) desde la secuencia analizada (Ye *y col.*, 2016). Esto se debe a que estos programas presentan en su flujo de trabajo la detección de puntos de quiebre durante el alineamiento de los *reads* al genoma de referencia, donde estos puntos estarían dando cuenta de la presencia de variantes complejas (Ivanov *y col.*, 2017). Este método de alineamiento con puntos de quiebre se basa en la división *reads*, donde primero se identifican *reads* con "extremos

discordantes" dado que un extremo mapean completamente con la secuencia de referencia y en el otro extremo no lo hacen; luego los extremos no mapeados de estos *reads* se agrupan o alinean mediante un ensamblaje *de novo*, permitiendo la identificación de *Indels* que son más largos que la longitud del r*ead,* pudiendo incluso determinar la presencia de variantes estructurales para esa región (Yang *y col.*, 2015). Por tal razón, estos programas identifican *Indels* con un amplio rango de tamaño dentro de las secuencias analizadas, generando imprecisiones al momento de analizar variaciones cortas en una región determinada del genoma, como se visualiza en el patrón de líneas verdes y azul del diagrama de Ventana Deslizable para el gen *MICA* (Figura 11). Adicionalmente, la identificación de *Indels* no validados en un grupo algoritmos para un determinado set de datos ya fue descrito para el algoritmo Pindel por Ghoneim *y col.*, 2014; donde se describe que para *Indels* de tamaño entre 1-5 pb este algoritmo genera un gran porcentaje de falsos positivos por cada variante validada por el grupo de programas utilizado.

Por otra parte, los algoritmos que presentaron mayor concordancia entre sí, Scalpel, SVC y Varscan, presentan un flujo de trabajo basado en alinear *reads* en una región fija del genoma de referencia y mediante un análisis con enfoque heurístico, otorgan puntajes al alineamiento de los *reads* contra la secuencia de referencia y seleccionan sólo aquellos que presentan un valor mayor (Hasan *y col.*, 2015). Cabe destacar, que los tres programas utilizan el *software* BWA (Burrows-Wheeler Aligner) para el alineamiento de secuencias, permitiendo la identificación de *Indels* cortos (<15 pb) desde secuencias de NGS (Li, 2012), y los resultados obtenidos para *MICA* concuerdan

con esto, dado que los 5 *Indels* detectados por estos programas no superaron las 11 pb de largo (Figura 12.A).

Al realizar un análisis en detalle de las 5 coordenadas que fueron comunes por los 5 algoritmos utilizados, se determinó que dos de ellas correspondían a variantes localizadas en regiones intrónica del gen *MICA* (Figura 12.A), las que no tendrían un efecto al momento de ser codificada la proteína como ha reportado Lin *y col.*, 2017.

De las 3 coordenadas restantes, chr6: 31380161 identificada en la base de datos genómicos Browser Beta y COSMIC (Figura 12.B y 12.C, respectivamente), se ha descrito como una variante del exón 5 del gen *MICA* que corresponde a una inserción de largo variable con un patrón de secuencia repetidas del triplete GCT, las que generan un desplazamiento del marco de lectura (*frameshift*) en la secuencia del gen (Figura 12.A). Estas inserciones con repeticiones del triplete GCT, que se traduce para el aminoácido Alanina (A), se localizan en el exón que codifica la región transmembrana (TM) de la proteína, y se han descrito alelos el gen *MICA* que presentan entre 4 a 10 repeticiones de GCT (A4, A5, A6, A7, A8, A9 y A10) (Fodil *y col.*, 1996), identificando que en el genoma de referencia se presentan 5 repeticiones (A5) (Mizuki *y col.*, 1997; Ota *y col.*, 1997). Adicionalmente se ha descrito la inserción de un nucleótido GCT > GGCT (posición rs67841474) en la región TM de las variantes A5.1 STR, la que genera un desplazamiento en el marco abierto de lectura, resultando en un codón de stop prematuro y codificando para una proteína MICA trunca en el dominio TM (Tamaki *y col.*, 2007), situación que se observa en las muestras tumorales de los pacientes CG6, CG14 y CG40, (Figuras 14, 16, 17 y 20).

Del análisis de los resultados para el gen *MICA*, se observa que se presentan inserciones en la posición chr6: 31380161 que estarían codificando una inserción de 1 a 4 Alaninas a la secuencia consenso del gen, variantes MICA-A6 y MICA-A9 respectivamente. Sin embargo, al comparar los resultados obtenidos entre los algoritmos, se tiene que para una misma muestra se identifican diferentes largos de secuencia para las repeticiones de GCT (Figura 13). De acuerdo a lo presentado en Zou *y col.,* 2006 estos resultados de análisis de NGS deben ser confirmados mediante un ensayo ortogonal de secuenciacion por Sanger para verificar la inserción de las repeticiones GCT detectadas en el exón 5. La confirmación del análisis de secuencias con patrones repetidos de tipo STR se debe a que el alineamiento de los *reads* puede generar imprecisiones en la identificación del largo del patrón de repeticiones para cada región analizada. Para ello se seleccionó un grupo de 5 muestras con un perfil particular de repeticiones de Alanina para ser analizadas por el método de Sanger (Figura 14). Donde por ejemplo, para la coordenada chr6: 31380161, las muestras CG-06 y CG-14 presentaron una alta concordancia en los algoritmos utilizados, dado que todos identifican una inserción de G/GCT (MICA-A6); sin embargo, en el caso de la muestra CG29 se observa una gran heterogeneidad en la identificaciones de *Indels* para esa posición, dado que mientras que SOAPindel y Pindel no identifican variantes en esa coordenada; y el resto de los algoritmos detectan inserciones de largo 2 (MICA-A6), 8 (MICA-A8) y 11 (MICA-A9) nucleótidos (Figura 14).

4.2 Confirmación de resultados de NGS mediante Sanger

A partir de los resultados obtenidos por la secuenciacion por Sanger, se tiene que los electroferogramas de las muestras analizadas presentan un perfil similar a lo descrito por Zou *y col.,* 2006. en el gen *MICA*, donde la inserción de repeticiones STR en la coordenada chr6: 31380161 genera un desplazamiento en la secuencia, observado como un doble *peak* en el electroferograma; esto es por una lectura en simultáneo de las secuencias que tienen la secuencia consenso del gen *MICA* (A5) y de aquellas secuencias que presentan la inserción. Por tal razón, al leer el electroferograma se deben identificar dos secuencias por separado, las que se inician en la bifurcación de la secuencia dado por el primer doble *peak.*

Los resultados observados dan cuenta de combinaciones de variantes MICA-STR para cada una de las muestras analizadas (Figura 16, 17, 18, 19 y 20); por ejemplo, para el análisis realizado al electroferograma de la muestra CG-06 y CG-14 se identifica que desde la bifurcación (base destacada en rojo, Figuras 16.D y 17.D), una de las secuencias corresponde a 6 repeticiones de GCT (MICA-A6), mientras que la otra hebra presenta 5 repeticiones de GCT junto con una inserción de G, lo que conlleva a un desplazamiento en el marco abierto de lectura (MICA-A5.1).

En el análisis de NGS, los algoritmos SVC y Scalpel detectaron inserciones de 8 nucleótidos en las muestras CG-29/33/40, sin embargo, estas secuencias no fueron detectadas mediante secuenciacion por Sanger. Tales resultados se consideran como falsos positivos por parte del análisis de NGS, dado que al revisar en detalle los

resultados de Sanger para esta mismas muestras (Figura 21), se tiene que estas muestras presentaron la variante MICA-A9 (inserción de 11 nucleótidos), por lo que la inserción de 8 nucleótidos detectada por NGS podría estar dando cuenta de un alineamiento local ambiguo, dada la presencia repeticiones de tipo STR en la región (Fang *y col.*, 2014).

Por otro lado, se tiene que los algoritmos Varscan y SOAP no estarían detectando determinadas variaciones en la secuencia de *MICA*, dado que de acuerdo a los resultados obtenidos por Sanger, las muestras CG-33 y 40 presentan el par A6/A9 y A5.1/A9, respectivamente; sin embargo, de acuerdo a los análisis de NGS estos algoritmos sólo detectaron la variante A9, por tal razón las variantes A6 y A5.1 serían falsos negativos (Figura 21) debido a la imposibilidad de ser detectados por estos programas (Hasan *y col.*, 2015). Adicionalmente, aunque el algoritmo Pindel detectó gran cantidad de *Indels* a través de la secuencia de *MICA* (sobre 300) (Tabla 4), este programa sólo detectó variantes en la posición chr6: 31380161 para la muestra CG-06. Esto se debe, a que este algoritmo no presenta buen rendimiento en la detección de *Indels* con secuencia repetidas en tándem (STR) (Ye *y col.*, 2009)

En general, en base a la validación de los resultados de NGS mediante secuenciación por Sanger se tiene que ningún algoritmo fue 100% certero al momento de detectar las repeticiones de GCT en la secuencia del gen *MICA* de las muestras analizadas. Sin embargo, es importante destacar que la detección de STR mediante NGS se ha descrito como un proceso complejo, dado que la presencia de secuencias repetidas en los *reads* conlleva a la introducción de ambigüedades en las posiciones del alineamiento, introduciendo en la mayoría de los casos falsos positivos en la identificación de *Indels*

(Narzisi & Schatz, 2015). Por tal razón y en base a los parámetros utilizados en el análisis del funcionamiento de cada algoritmo (Figuras 22 y 23), se propone que para aumentar la precisión en la detección de *Indels* mediante NGS en la secuencia de *MICA* se utilicen a lo menos 3 de los programas analizados; dado que de acuerdo a los resultados de esta tesis, la detección de un *Indel* por este número mínimo de programas (Figura 14) se correlaciona en un 100% con su detección por Sanger (Figura 21).

En particular, de los 3 programas a utilizar, se deben considerar como base a los algoritmos Scalpel y SVC, debido a que se presentan como los algoritmos con mejor sensibilidad (Tabla 5) en la detección de STR en la secuencia de MICA, dado detectaron en general, todas las variantes MICA-STR obtenidas mediante Sanger en las muestras analizadas, y adicionalmente detectaron la inserción en la posición contigua chr6: 31380160 T/TG, que estaría dando luces de la identificación de la variante MICA-A5.1 mediante análisis de NGS en las secuencias analizadas. No obstante, dada su precisión media, se sugiere considerar un tercer algoritmo de los analizados con el objetivo de realizar una identificación de *Indels* con altos parámetros de rellamado y precisión (Figura 23)

4.3 *Indels* en el gen *MICA* y su efecto en la proteína

Con el objetivo de analizar *in silico* el efecto de los *Indels* detectados en la proteína MICA, se utilizó la base de datos SIFT que contiene información de variantes codificantes no sinónimas y analiza el efecto fenotípico de tales variantes en genes humanos. La versión SIFTIndel, incluye el efecto de *Indels* en la estabilidad de proteínas

(Kumar *y col.*, 2009). En particular, se obtuvo que las inserciones de G (MICA-A5.1), CT (MICA-A6) y CTGCTGCTGCT (MICA-A9) en la coordenada chr6: 31380161 tendrían un efecto neutral en la proteína, dado por un valor SIFT de 0.787 (menor a 0.05 se considera como un cambio significativo para la proteína) (Figura 24)

Sin embargo, a pesar de su efecto neutral, la inserción de G en la posición chr6: 31380161 (MICA A5.1) estaría generando un desplazamiento del marco de lectura en la secuencia del gen. En la literatura se ha reportado que el alelo de MICA que contiene la secuencia A5.1, MICA*008, presenta la liberación de la proteína a través de exosomas desde la superficie de la célula que la expresa, esto se debe a que se agrega una tallo GPI a la secuencia del dominio TM truncado de MICA (Mizuki *y col*, 1997). La liberación de MICA* 008 como una proteína multimérica unida a la membrana, al igual que la proteína soluble monomérica liberada por *shedding* enzimático (Waldhauer *y col.*, 2008), la hace un regulador negativo potente de NKG2D y provoca la pérdida de la respuesta citolítica de las células NK (Ashiru *y col.*, 2010). Adicionalmente, el polimorfismo MICA-A5.1 se ha asociado con enfermedades autoinmunes como diabetes tipo I (Triolo *y col.*, 2009), infecciones por Citomegalovirus (Moenkemeyer y *col.*, 2009) y varias neoplasias (Chen & Gyllensten, 2014), apoyando así su papel en la respuesta inmune y el desarrollo tumoral. En el presente estudio, para las muestras que presentaron mediante NGS la inserción T/TG en la posición chr6: 31380161(MICA-A5.1) (CG-06/14/40) (Figura 14), sus resultados fueron validados mediante secuenciación por Sanger (Figura 21).

Por otro lado, la inserción de once nucleótidos CTGCTGCTGCT en la posición chr6: 31380161 (MICA-A9), genera un desplazamiento en el marco de lectura y adiciona 4 alaninas en la secuencia de la proteina. Y se ha descrito que la variante MICA-A9 se asocia con un mayor riesgo de GC en un estudio caso-control en población taiwanesa (Lo *y col.,* 2004) y un mayor riego de carcinoma hepatocelular inducido por virus de hepatitis B; detectando niveles de sMICA significativamente más altos con esta variante de microsatélite en el suero de los pacientes (Tong *y col.*, 2013). En la presente tesis, para las muestras que presentaron mediante el análisis de NGS la inserción G/GCTGCTGCTGCT en la posición chr6: 31380161 (MICA-A9)(CG-29/33/40) (Figura 14), sus resultados fueron validados en su totalidad mediante secuenciación por Sanger (Figura 21). Al extrapolar estos resultados al resto de las muestras analizadas por NGS, se identifican 17 muestras que presentan la inserción de 11 nucleótidos en chr6: 31380161 detectada por al menos 3 programadas, lo que podría indicar la presencia de la variante MICA-A9 en estas muestras (Tablas II y III. Anexo 3). Permitiendo inferir que 20 de las 50 muestras analizadas presentan la variante MICA-A9.

Ante el gran número de muestras que presenta la variante A9 del microsatélite, se sugiere que al combinarla con variantes del ectodominio del ligando MICA, se podría favorecer la liberación de la molécula soluble en pacientes con CG y en este punto, sería interesante evaluar la frecuencia de esta variante en la población chilena. En particular, se propone identificar la presencia de estas mutaciones en el genoma de pacientes chilenos sanos, con el fin de determinar si estas variantes generan predisposición a generar un tumor.

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 son liberados mediante exosomas 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 & Held, 2006).

Para el resto de las muestra analizadas por NGS, se detectaron inserciones de largo 2 y 8 nucleótidos en la coordenada chr6: 31380161, correspondiendo a las variantes MICA-A6 y MICA-A8, respectivamente (Tabla IV, V y VI. Anexo 3); sin embargo, este resultado no es concluyente debido a que por una parte tales variantes no fueron detectadas por más de 3 programas para cada muestra analizadas, y adicionalmente, tales *Indels* no fueron revisados por Sanger para ser validados. No obstante, dada la complejidad en la detección de variantes de tipo STR en el gen *MICA,* los programas utilizados obtuvieron *Indels* con un largo descrito para inserciones en el gen *MICA,* indicando que la identificación de *Indels* no es un proceso al azar.

Finalmente, los resultados obtenidos en este estudio nos permiten corroborar la hipótesis propuesta. Donde la utilización de un conjunto de algoritmos de alineamiento local permitió la detección de variaciones genéticas de tipo *Indel* en la secuencia del gen *MICA*, cuyo efecto podría estar relacionado con la liberación de ligandos solubles y la progresión tumoral. Sin embargo, se requieren análisis más exhaustivos para poder determinar el real efecto de las variantes detectadas y su efecto en la conformación de la proteína.

Ante esto es importante destacar que si bien *a priori* se podría pensar que los *Indels* 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 la liberación de ligandos solubles desde la célula que los expresa (Obuchi *y col.*, 2001).

En la actualidad la identificación de variaciones de tipo *Indels* desde datos de secuenciación masiva es un reto debido a la dificultad de distinguir los verdaderos acontecimientos de artefactos derivados de la PCR y errores de alineamiento, entre otros. Por tal razón, en base a los resultados de la presente tesis, y dada la amplia gama de algoritmos disponibles para la detección de este tipo de variantes es importante identificar las características del *Indel* que se estudia, donde por ejemplo para *Indels* complejos y de largo considerable (> 1000pb) se sugiere utilizar Pindel (Yang *y cols.*, 2015), mientras que para *Indels* pequeños (<15pb) que presenten repeticiones de tipo STR en su secuencias, se sugiere la utilización de Scalpel o SVC (Narzisi & Schatz, 2015). No obstante, en ambas situaciones los resultados obtenidos se deben corroborar mediante secuenciación por Sanger. En nuestro caso, se sugiere la combinación de Scalpel junto a SVC, dado que aún cuando no se logró el 100% de concordancia entre ambos programas sus resultados fueron comparativamente bastantes cercanos para *Indels* detectados en la secuencia de *MICA* mediante Sanger.

Por otra parte, la detección de posibles haplotipos en el gen *MICA,* dados por la combinación de variantes MICA-STR, A5, A5.1, A6 y A9 en las muestras en estudio, se podría correlacionar con que este gen está 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 variantes 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.

Los resultados obtenidos en la presente Tesis nos permiten postular que la utilización de diversos algoritmos permiten detectar mutaciones de tipo *Indels* en la secuencia del gen *MICA*. Las que en relación a las muestras analizadas, generan variaciones con efecto neutro en la proteína; sin embargo generan un desplazamiento en el marco abierto de lectura, alterando la región transmembrana de la proteína y promoviendo posiblemente la liberación soluble del ligando MICA, y que esto podría estar dando cuenta de 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.

CONCLUSIONES

- Los resultados de nuestro estudio demuestran que es posible detectar mutaciones de tipo *Indels* en el gen *MICA* a partir del análisis de secuenciación masiva del ADN genómico proveniente de pacientes con adenocarcinoma gástrico.
- En base al grupo de algoritmos utilizados, se observó heterogeneidad tanto en el número como en el largo de los *Indels* detectados a través de la secuencia del gen *MICA*
- El análisis bioinformático para el gen *MICA* permitió determinar que dada la presencia de *Indels* de tipo STR, se sugiere utilizar en conjunto los algoritmos SVC, Scalpel y un tercero de los analizados, debido a que de acuerdo al protocolo desarrollado, se aumentaría la precision en la detección de este tipo de variante.
- Con el protocolo propuesto en esta tesis, se determinó que una proporción cercana del 40% de los pacientes analizados presentaron la inserción que codifica para MICA-A9, variante que se ha asociado con un mayor riesgo de CG en población Taiwanesa.
- En base a los resultados obtenidos en este trabajo, se comprobó la hipótesis que plantea que la utilización de un conjunto de algoritmos de alineamiento local

permiten la detección de *Indels* en la secuencia del gen *MICA*, cuyo efecto podría estar relacionado con la liberación de ligandos solubles y la progresión tumoral. No obstante, se requieren de análisis adicionales para determinar el efecto de estas variantes en la conformación de la proteína MICA.

- Futuros estudios podrían centrarse en la detección de haplotipos en la secuencias del gen *MICA,* dados por la combinación de variantes MICA-A5.1 y MICA-A9, de modo de analizar su función como posibles marcadores moleculares o predictores de riesgo a padecer CG.

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.

Ashiru, O., Boutet, P., Fernandez-Messina, L., Aguera-Gonzalez, S., Skepper, J., Vales-Gomez, M., Reyburn, H. 2010. Natural killer cell cytotoxicity is suppressed by exposure to the human NKG2D ligand MICA*008 that is shed by tumor cells in exosomes. Cancer Research. 70: 481–489.

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.

Bansal, K. 2012. Analysis of Sliding Window Protocol for Connected Node. International Journal of Soft Computing and Engineering. 2: 292-294.

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.

Breder, V., Baranova, A., Chernenko, P., Ivanov, M., Laktionov, K., Musienko, S., Mileyko, V., Novikova, E., Telysheva, E. 2017. Towards standardization of next-generation sequencing of FFPE samples for clinical oncology: intrinsic obstacles and possible solutions. Journal of Translational Medicine. 15: 22.

Burnet, F. M. 1971. Immunological Surveillance in Neoplasia. Immunological Reviews. 7: 3–25.

Caligiuri, M. 2008. Human natural killer cells. Blood. 112: 461-469.

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.

Champsaur, M. & Lanier, L. 2010. Effect of NKG2D ligand expression on host immune responses. Immunological Reviews. 235: 267–285.

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.

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 inmunity. Seminars in Cancer Biology. 16: 333-343.

Coudert, J., Scarpellino, L., Gros, F., Vivier, E., Held, W. 2008. Sustained NKG2D engagement induces cross-tolerance of multiple distinct NK cell activation pathways. Blood. 111: 3571-3578.

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.

DePristo, M., Banks, E., Poplin, R., Garimella, K., Maguire, J., Hartl, C., Daly, M. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics. 43: 491–498.

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.

Diefenbach, A. & Raulet, D. H. 2002. The innate immune response to tumors and its role in the induction of T-cell immunity. Immunological Reviews, 188: 9–21.

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.

Fang, H., Wu, Y., Narzisi, G., O'Rawe, J. A., Barrón, L. T., Rosenbaum, J., Lyon, G. 2014. Reducing INDEL calling errors in whole genome and exome sequencing data. Genome Medicine. 6: 89.

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.

Fernández-Messina, L., Reyburn, H. T., Valés-Gómez, M. 2012. Human NKG2D-ligands: cell biology strategies to ensure immune recognition. Frontiers in Immunology. 3: 299.

Fodil, N., Laloux, L., Wanner, V. 1996. Allelic repertoire of the human MHC class I MICA gene. Immunogenetics. 44:351–357.

Frigoul, A. & Lefranc, M. 2005. MICA: Standardized IMGT allele nomenclature, polymorphisms and diseases. Recent Research Developments in Human Genetics. 3: 95-145.

Gárate-Calderón, V. 2016. Análisis de variaciones genéticas del ligando MICA en tumores de pacientes con adenocarcinoma gástrico. Seminario de título para

obtener el título de Ingeniera en Biotecnología Molecular. Universidad de Chile. Tutor: Dra. María Carmen Molina, Co-Tutor: Dr. Ricardo Armisen.

Gárate-Calderón, V., Gutierrez, M., Morales, M. & Molina, M. C. (en revisión) Detection of SNPs in *MICA* in Chilean gastric cancer patients using Next-Generation Sequencing.

Garrido-Tapia, M., Hernandez, C., Ascui, G., Kramm, K., Morales, M., Garate, V., Zuniga, R., Bustamante, M., Aguillon, J.C., Catala, N.D., Ribeiro, C.H., Molina, M.C. 2017. STAT3 inhibition by STA21 increases cell surface expression of MICB and the release of soluble MICB by gastric adenocarcinoma cells. Immunobiology. 222:1043-1051.

Ghadially, H., Brown, L., Lloyd, C., Lewis, L., Lewis, A., Dillon, J., Wilkinson, R. W. 2017. MHC class I chain-related protein A and B (MICA and MICB) are predominantly expressed intracellularly in tumour and normal tissue. British Journal of Cancer.116: 1208–1217.

Ghoneim, D. H., Myers, J., Tuttle, E., Paciorkowski, A. 2014. Comparison of insertion/deletion calling algorithms on human next-generation sequencing data. BMC Research Notes. 7: 864.

Gonzalez, S., Groh, V., Spies, T. 2006. Immunobiology of human NKG2D and its ligands. Current Topics in Microbiology and Immunology. 298: 121–138.

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.

Hasan, M, Wu, X., Zhang, L. 2015. Performance evaluation of indel calling tools using real short-read data. Human Genomics. 9: 20.

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: 1017-1020.

Hwang, S., Kim, E., Lee, I., Marcotte, E. M. 2015. Systematic comparison of variant

calling pipelines using gold standard personal exome variants. Scientific Reports. 5: 17875.

Iengar, P. 2012. An analysis of substitution, deletion and insertion mutations in cancer genes. Nucleic Acids Research. 40: 6401–6413.

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.

Ivanov, M., Laktionov, K., Breder, V., Chernenko, P., Novikova, E., Telysheva, E. 2017. Towards standardization of next-generation sequencing of FFPE samples for clinical oncology: intrinsic obstacles and possible solutions. Journal of Translational Medicine. 15: 22.

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. (Reprinted from Nature 319: 675-678 (1986)). Journal of Immunology. 174: 6566-6569.

Koboldt, D., Zhang, Q., Larson, D., Shen, D., McLellan, M., Lin L. 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Research. 22: 568–576.

Korbel, O., Urban, A., Grubert, F., Du, J., Royce, T. E., Starr, P., Gerstein, M. 2007. Systematic prediction and validation of breakpoints associated with copy-number variants in the human genome. Proceedings of the National Academy of Sciences of the United States of America. 104: 10110–10115.

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., Hau 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.

Lamb, A. & Chen, L.-F. 2013. Role of the Helicobacter pylori-induced inflammatory response in the development of gastric cancer. Journal of Cellular Biochemistry. 114: 491–497.

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 Infectous Disseases. 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 cáncer susceptibility biomarkers. Revista Médica de Chile. 134: 1107-1115.

Li, H. 2012. Exploring single-sample SNP and INDEL calling with whole-genome de novo assembly. Bioinformatics. 28: 1838–1844.

Li, H. 2014. Toward better understanding of artifacts in variant calling from high-coverage samples. Bioinformatics. 30: 2843–2851.

Li, S., Li, R., Li, H., Lu, J., Li, Y., Bolund, L., Wang, J. 2013. SOAPindel: Efficient identification of indels from short paired reads. Genome Research. 23: 195–200.

Lo, S.-S., Lee, Y.-J., Wu, C.-W., Liu, C.-J., Huang, J.-W., & Lui, W.-Y. 2004. The increase of MICA gene A9 allele associated with gastric cancer and less schirrous change. British Journal of Cancer. 90: 1809–1813.

Lin, M., Whitmire, S., Chen, J., Farrel, A. 2017. Effects of short indels on protein structure and function in human genomes. Scientific Reports 7: 9313.

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.

Marcus, A., Gowen, B. G., Thompson, T. W., Iannello, A., Ardolino, M., Deng, W., Raulet, D. H. 2014. Recognition of tumors by the innate immune system and natural killer cells. Advances in Immunology.122: 91–128.

Martinez-Chamorro, A., Moreno, A., Gomez-Garcia, M. 2016. MICA*A4 protects against ulcerative colitis whereas MICA*A5.1 is associated with abscesses formation and the age on onset. Clinical and Experimental Immunology. 184: 323-331.

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.

Meldrum, C., Doyle, M. A., Tothill, R. W. 2011. Next-Generation Sequencing for Cancer Diagnostics: a Practical Perspective. The Clinical Biochemist Reviews. 32: 177–195.

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 thetransmembrane 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.

Moenkemeyer, M., Heiken, H., Schmidt, R.E., Witte, T. 2009. Higher risk of cytomegalovirus reactivation in human immunodeficiency virus-1-infected patients homozygous for MICA5.1. Hum Immunol.70:175–178

Narzisi, G., O'Rawe, J., Iossifov, I., Fang, H., Lee, Y.H., Wang, Z. 2014. Accurate de novo and transmitted indel detection in exome-capture data using microassembly. Nature Methods. 11:1033–1036.

Narzisi, G. & Schatz, M. 2015. The Challenge of Small-Scale Repeats for Indel Discovery. Frontiers in bioengineering and biotechnology. 3: 8.

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.

Ota, M., Katsuyama, Y., Misuki, N., Ando, H., Furikata, K., Ono, S., Pivetti-Pezzi, P., Tabbra, K.F., Palimeris, G. 1997. Trinucleotide repeat polymorphism with exon 5 of the MICA gene (MHC class I chain-related gene A): allele frequency data in the nine population groups Japanese, Northern Han, Hui, Uygur, Kazakhstan, Iranian, Saudi Arabian, Greek and Itlain. Tissue Antigens. 49:448–454.

Podack, E. & Dennert, G. 1983. Assembly of two types of tubules with putative cytolytic function by cloned natural killer cells. Nature. 302: 442-445.

Ratan, A., Olson, T., Loughran, T., Miller, W. 2015. Identification of indels in next-generation sequencing data. BMC Bioinformatics. 16: 42.

Ribeiro, C. H., Kramm, K., Gálvez-Jirón, F., Pola, V., Bustamante, M., Contreras, H. R., Molina, M. C. 2016. Clinical significance of tumor expression of major histocompatibility complex class I-related chains A and B (MICA/B) in gastric cancer patients. Oncology Reports. 35: 1309–1317.

Robinson, J., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E., Getz, G., Mesirov, J. 2011. Integrative Genomics Viewer. Nature Biotechnology. 29: 24–26.

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.

Schröder, J., Hsu, A., Boyle, S., Macintyre, G., Cmero, M., Tothill, R. 2014. Socrates: identification of genomic rearrangements in tumour genomes by re-aligning soft clipped reads. Bioinformatics. 30: 1064–1072.

Serrano, A., Menares-Castillo, E., Garrido-Tapia, M., Ribeiro, C.H., Hernandez, C., Mendoza-Naranjo, A., Gatica-Andrades, M., Valenzuela-Diaz, R., Zuniga, R., Lopez, M.N., Salazar-Onfray, F., Aguillon, J.C., Molina, M.C. 2011. Interleukin 10 decreases MICA expression on melanoma cell surface. Immunol Cell Biol. 89: 447-457.

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.

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.

Tattini, L., D'Aurizio, R., Magi, A. 2015. Detection of Genomic Structural Variants from Next-Generation Sequencing Data. Frontiers in Bioengineering and Biotechnology. 3: 92.

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.

Tong, H.V., Toan, N.L., Song, L.H., Bock, C.T., Kremsner, P.G., Velavan, T.P. 2013 Hepatitis B virus-induced hepatocellular carcinoma: functional roles of MICA variants. Journal of Viral Hepatitis. 20: 687–698.

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.

Triolo, T.M., Baschal, E., Armstrong, T., Toews, C.S., Fain, P.R., Rewers, M., Yu, L., Miao, D., Eisenbarth, G., Gottlieb, P., Barker, J. 2009. Homozygosity of the polymorphism MICA5.1 identifies extreme risk of progression to overt adrenal insufficiency among 21-hydroxylase antibody-positive patients with type 1 diabetes. Journal of Clinical Endocrinology and Metabolism. 94: 4517–4523.

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.

Waldhauer, I., Goehlsdorf, D., Gieseke, F., Weinschenk, T., Wittenbrink, M., Ludwig, A., Stevanovic, S., Rammensee, H.G., Steinle, A. 2008. Tumor-associated MICA is shed by ADAM proteases. Cancer Research. 68: 6368–6376.

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.

Xue, Y., Lameijer, E., Ye, K., Zhang, K., Chang, S., Wang, X., Yu, J. 2016. Precision Medicine: What Challenges Are We Facing? Genomics, Proteomics & Bioinformatics. 14: 253–261.

Yang, R., Nelson, A., Henzler, C., Thyagarajan, B., Silverstein, K. 2015. ScanIndel: A hybrid framework for indel detection via gapped alignment, split reads and de novo assembly. Genome Medicine. 7: 127.

Ye, K., Schulz M., Long Q., Apweiler R., Ning Z. 2009 Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics. 25: 2865-71.

Ye, K., Wang, J., Jayasinghe, R., Lameijer, E., McMichael, J.F., Ning, J. 2016. Systematic discovery of complex insertions and deletions in human cancers. Nature Medine. 22:97–104.

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: 2375-2380.

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.

Zou, Y., Han, M., Wang, Z., Stastny, P. 2006. MICA allele-level typing by sequence-based typing with computerized assignment of polymorphic sites and short tandem repeats within the transmembrane region. Human Immunology. 67: 145-151.

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.

ANEXOS

Anexo 1. Consentimiento Informado.

UNIVERSIDAD DE CHILE

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: pgonzalez@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 alergica 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.

3 . DIC 2011

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. Acta de aprobación del Comité de Ética.

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

1/2

31 MAYO 2011

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 Biagini 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 MAY 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: si
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

Anexo 3. Resultados de NGS para el resto de las muestras en estudio.

Tabla I. Anexo 3. Resumen del análisis NGS para muestras que presentan la variante MICA-A5.1

Muestra	Programa	Cromosoma	Coordenada	Gen	Ref	Alt	Variante	Largo	Alelo MICA	Efecto	Posición
CG04	Scalpel	chr6	31380160	MICA	T	TG	INSERCIÓN	1	A5.1	frameshift_variant	p.Gly318fs
	SOAP	chr6	31380160	MICA	TG	TGG	INSERCIÓN	1		frameshift_variant	p.Gly318fs
	SVC	chr6	31380160	MICA	TG	TGG	INSERCIÓN	1		frameshift_variant	p.Gly318fs
	Varscan	chr6	31380160	MICA	T	TG	INSERCIÓN	1		frameshift_variant	p.Gly318fs
CG16	Scalpel	chr6	31380160	MICA	T	TG	INSERCIÓN	1	A5.1	frameshift_variant	p.Gly318fs
	SOAP	chr6	31380160	MICA	TG	TGG	INSERCIÓN	1		frameshift_variant	p.Gly318fs
	SVC	chr6	31380160	MICA	TG	TGG	INSERCIÓN	1		frameshift_variant	p.Gly318fs
	Varscan	chr6	31380160	MICA	T	TG	INSERCIÓN	1		frameshift_variant	p.Gly318fs
	Pindel	chr6	31380160	MICA	T	TG	INSERCIÓN	1		frameshift_variant	p.Gly318fs
CG34	Scalpel	chr6	31380160	MICA	T	TG	INSERCIÓN	1	A5.1	frameshift_variant	p.Gly318fs
	SVC	chr6	31380160	MICA	T	TG	INSERCIÓN	1		frameshift_variant	p.Gly318fs
	Varscan	chr6	31380160	MICA	T	TG	INSERCIÓN	1		frameshift_variant	p.Gly318fs
CG35	Scalpel	chr6	31380160	MICA	T	TG	INSERCIÓN	1	A5.1	frameshift_variant	p.Gly318fs
	SOAP	chr6	31380160	MICA	TG	TGG	INSERCIÓN	1		frameshift_variant	p.Gly318fs
	SVC	chr6	31380160	MICA	T	TG	INSERCIÓN	1		frameshift_variant	p.Gly318fs
	Varscan	chr6	31380160	MICA	T	TG	INSERCIÓN	1		frameshift_variant	p.Gly318fs
CG38	Scalpel	chr6	31380160	MICA	T	TG	INSERCIÓN	1	A5.1	frameshift_variant	p.Gly318fs
	SOAP	chr6	31380160	MICA	TG	TGG	INSERCIÓN	1		frameshift_variant	p.Gly318fs
	SVC	chr6	31380160	MICA	T	TG	INSERCIÓN	1		frameshift_variant	p.Gly318fs
	Varscan	chr6	31380160	MICA	T	TG	INSERCIÓN	1		frameshift_variant	p.Gly318fs
CG42	Scalpel	chr6	31380160	MICA	T	TG	INSERCIÓN	1	A5.1	frameshift_variant	p.Gly318fs
	SOAP	chr6	31380160	MICA	TG	TGG	INSERCIÓN	1		frameshift_variant	p.Gly318fs
	SVC	chr6	31380160	MICA	T	TG	INSERCIÓN	1		frameshift_variant	p.Gly318fs
	Varscan	chr6	31380160	MICA	T	TG	INSERCIÓN	1		frameshift_variant	p.Gly318fs

Tabla II. Anexo 3. Resumen del análisis NGS para muestras que presentan la variante MICA-A9

	Programa	Cromosoma	Coordenada	Gen	Ref	Alt	Variante	Largo	Alelo MICA	Efecto	Posición
CG08	Scalpel	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11	A9	frameshift_variant	p.Gly318fs
	SOAP	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
	SVC	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
	Varscan	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
CG13	Scalpel	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11	A9	frameshift_variant	p.Gly318fs
	SOAP	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
	SVC	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
CG16	Scalpel	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11	A9	frameshift_variant	p.Gly318fs
	SVC	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
	Varscan	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
CG19	Scalpel	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11	A9	frameshift_variant	p.Gly318fs
	SOAP	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
	SVC	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
CG24	Scalpel	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11	A9	frameshift_variant	p.Gly318fs
	SOAP	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
	SVC	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
CG27	Scalpel	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11	A9	frameshift_variant	p.Gly318fs
	SOAP	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
	SVC	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
	Varscan	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
CG28	Scalpel	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11	A9	frameshift_variant	p.Gly318fs
	SOAP	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
	SVC	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
CG34	Scalpel	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11	A9	frameshift_variant	p.Gly318fs
	SOAP	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
	SVC	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs

Tabla III. Anexo 3. Resumen del análisis NGS para muestras que presentan la variante MICA-A9

	Programa	Cromosoma	Coordenada	Gen	Ref	Alt	Variante	Largo	Alelo MICA	Efecto	Posición
CG36	Scalpel	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11	A9	frameshift_variant	p.Gly318fs
	SOAP	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
	SVC	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
	Varscan	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
CG37	Scalpel	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11	A9	frameshift_variant	p.Gly318fs
	SOAP	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
	SVC	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
CG38	Scalpel	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11	A9	frameshift_variant	p.Gly318fs
	SVC	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
	Varscan	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
CG44	Scalpel	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11	A9	frameshift_variant	p.Gly318fs
	SVC	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
	Varscan	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
CG45	Scalpel	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11	A9	frameshift_variant	p.Gly318fs
	SOAP	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
	SVC	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
CG49	Scalpel	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11	A9	frameshift_variant	p.Gly318fs
	SOAP	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
	SVC	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
	Varscan	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
CG51	Scalpel	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11	A9	frameshift_variant	p.Gly318fs
	SOAP	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
	SVC	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
CG52	Scalpel	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11	A9	frameshift_variant	p.Gly318fs
	SOAP	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
	SVC	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
	Varscan	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
CG53	Scalpel	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11	A9	frameshift_variant	p.Gly318fs
	SOAP	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs
	SVC	chr6	31380161	MICA	G	GCTGCTGCTGCT	INSERCIÓN	11		frameshift_variant	p.Gly318fs

Tabla IV. Anexo 3. Resumen del análisis NGS para muestras que presentan inserción de largo 2 y 8 nucleótidos en la coordenada chr6: 31380161.

soap	CG01	chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
scalpel	CG02	chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
scalpel	CG06	chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
soap		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
varscan		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc	CG08	chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
scalpel	CG13	chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
scalpel	CG14	chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
soap		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
varscan		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc	CG15	chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
svc	CG16	chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
scalpel	CG18	chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
varscan		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
scalpel	CG19	chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
scalpel	CG21	chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
scalpel		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
soap	CG22	chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
scalpel	CG23	chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
scalpel	CG24	chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
scalpel		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH

Tabla V. Anexo 3. Resumen del análisis NGS para muestras que presentan inserción de largo 2 y 8 nucleótidos en la coordenada chr6: 31380161.

scalpel	CG25	chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
soap		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
varscan		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc	CG27	chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
scalpel	CG28	chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
scalpel	CG29	chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
scalpel	CG30	chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
scalpel	CG32	chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
scalpel	CG33	chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
scalpel		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
scalpel	CG34	chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
scalpel	CG35	chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc	CG36	chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
scalpel	CG37	chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
scalpel	CG38	chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
scalpel	CG39	chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
scalpel	CG40	chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
scalpel	CG41	chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
varscan		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH

Tabla VI. Anexo 3. Resumen del análisis NGS para muestras que presentan inserción de largo 2 y 8 nucleótidos en la coordenada chr6: 31380161.

svc	CG42	chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
scalpel	CG43	chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
scalpel		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
soap		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
scalpel	CG44	chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
scalpel	CG45	chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
varscan		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc	CG49	chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
scalpel	CG50	chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
scalpel		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
scalpel	CG51	chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
svc	CG52	chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
scalpel	CG53	chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
scalpel		chr6	31380161	MICA	G	GCTGCT	INS	5	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH
scalpel	CG57	chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
soap		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
scalpel	CG58	chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
soap		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
scalpel	CG59	chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCT	INS	2	frameshift_variant	p.Gly318fs	HIGH
svc		chr6	31380161	MICA	G	GCTGCTGCT	INS	8	frameshift_variant	p.Gly318fs	HIGH