

**UNIVERSIDAD DE CHILE
FACULTAD DE MEDICINA
ESCUELA DE POSTGRADO**



**IDENTIFICACIÓN DE VARIANTES EN EL PROMOTOR DE
RAET1E QUE ORIGINEN CAMBIOS EN SITIOS DE UNIÓN A
FACTORES DE TRANSCRIPCIÓN, EN TUMORES DE
PACIENTES CON ADENOCARCINOMA GÁSTRICO**

MARCELA ALEJANDRA MORALES GEDDA

**TESIS PARA OPTAR AL GRADO DE MAGISTER EN CIENCIAS BIOMÉDICAS
MENCIÓN INMUNOLOGÍA**

Director de Tesis: Prof. Dra. María Carmen Molina Sampayo

Co-Director de Tesis: Prof. Dr. Ricardo Armisen Yañez

2016

**UNIVERSIDAD DE CHILE
FACULTAD DE MEDICINA
ESCUELA DE POSTGRADO**

INFORME DE APROBACION TESIS DE MAGÍSTER

Se informa a la Comisión de Grados Académicos de la Facultad de Medicina, que la Tesis de Magister presentada por la candidata

MARCELA ALEJANDRA MORALES GEDDA

ha sido aprobada por la Comisión Informante de Tesis como requisito para optar al Grado de **Magister en Ciencias Biomédicas**, con **mención en Inmunología** en el Examen de Defensa de Tesis rendido el día 25 de abril de 2016.

Prof. Dra. María Carmen Molina

Directora de Tesis

Prof. Dr. Ricardo Armisen

Co-Director de Tesis

Programa Disciplinario de Inmunología, ICBM, Facultad de Medicina, Universidad de Chile.

COMISION INFORMANTE DE TESIS

Prof. Dr. Patricio González.

Prof. Dra. Carolina Ribeiro.

Prof. Dra. Daniela Sauma.

Prof. Dr. Diego Catalán.

Presidente Comisión

A mi padre, Patricio Morales Gagliardi, quien pese a nunca comprender mucho lo que su hija hacía me apoyó. A mis hermanos, Toño, Cris y Lore, quienes me apodaron “ratona de biblioteca”. Y en especial a ti, Carmen Gedda Ortiz, madre y amiga, quien muy atenta escuchaste el relato de mis travesías en el lab. Los adoro.

AGRADECIMIENTOS

Son muchas las personas que conocí en esta etapa de mi vida, y a las cuales agradezco inmensamente porque ayudaron a nutrir este camino de diversas formas. En primer lugar quiero agradecer a mi tutora, María Carmen Molina, que permitió mi incorporación en su laboratorio sin mayores antecedentes más que mi desempeño en el magíster de inmunología. Abrió sus puertas entregándome la confianza e independencia para desarrollar esta tesis. Cómo no agradecer a cada uno de los ex y actuales integrantes del laboratorio de inmunovigilancia y evasión inmune por su ayuda y buenos momentos: *Caro, Maca, Kari, Agnes, Daniel, Lorenzo, Felipe, Leito, Gabriel, Valentina, Jaime y Bastián.*

Quisiera agradecer por conocer y compartir con Carolina Ribeiro, quien siempre tuvo palabras de motivación y apoyo en este complejo proceso.

Durante el desarrollo de mi tesis de magister pude establecer estrechos lazos con otros laboratorios de la facultad de Medicina. Por lo tanto, quisiera agradecer a mi co-tutor Ricardo Armisén, quien apoyó activamente el desarrollo de esta tesis, tanto en lo personal como en lo profesional. Quisiera también agradecer a cada uno de los integrantes de su equipo, en especial a *Dani, Marisel, Eduardo, la Dra. Marcelain y muy especialmente a mi amiga Jessica Toro*, por toda la ayuda prestada en el desarrollo y avance de esta investigación.

No puedo dejar fuera de este agradecimiento de cada uno de los integrantes del Programa de Inmunología, en especial a *Javier, Oriana, Inelia y Juani.*

Por último agradezco a mi pareja, quien apoyó incondicionalmente este proceso, y toleró mis cambios de ánimos, el estrés y mis pérdidas de cordura, entregándome confianza, paz y nuevas aventuras.

Sinceramente, a todos ellos, muchísimas gracias.

INDICE DE CONTENIDOS

INDICE DE CONTENIDOS	2
ÍNDICE DE FIGURAS	7
ÍNDICE DE TABLAS.....	8
ABREVIATURAS.....	9
RESUMEN.....	10
ABSTRACT.....	13
INTRODUCCIÓN.....	15
1. CÁNCER GÁSTRICO.....	15
2. INMUNIDAD ANTITUMORAL Y NKG2D	18
3. LIGANDOS DEL RECEPTOR DE ACTIVACIÓN NKG2D.....	19
4. EXPRESIÓN DE LOS LIGANDOS DE NKG2D.....	22
5. REGULACIÓN DE LA EXPRESIÓN DE LOS LIGANDOS DE NKG2D.....	23
5.1 Regulación Transcripcional	25
6. MECANISMOS DE EVASIÓN MEDIADOS POR EL SISTEMA NKG2D Y SUS LIGANDOS.....	27
7. VARIABILIDAD DE LOS NKG2DLs.....	27
7.1 MICA Y MICB	27
7.2 RAET1s (ULBPs).....	29
8. RAET1E (ULBP4).....	31
HIPÓTESIS.....	35
OBJETIVOS.....	36
1. OBJETIVO GENERAL.....	36
2. OBJETIVOS ESPECÍFICOS.....	36
METODOLOGÍA.....	37
1. PACIENTES Y MUESTRAS BIOLÓGICAS	37
2. OBTENCIÓN DE ADN GENÓMICO	37
3. ANÁLISIS CUANTITATIVO Y CUALITATIVO DEL ADN GENÓMICO	38
3.1 Electroforesis en gel de Agarosa.....	38
3.2 Cuantificación del ADN genómico.....	38
3.3 Electroforesis Capilar Paralela	40
4. SECUENCIACIÓN DEL PROMOTOR DE RAET1E	41
4.1 Diseño del proyecto de Secuenciación del promotor de RAET1E	41
4.2 Preparación Librerías Genéticas.....	41
4.3 Secuenciación masiva del promotor de RAET1E	44

5.	ANÁLISIS BIOINFORMÁTICO POST-SECUENCIACIÓN E IDENTIFICACIÓN DE VARIACIONES EN EL PROMOTOR DE RAET1E.....	45
6.	ESTUDIO <i>IN SILICO</i> DE LA REGION DEL PROMOTOR DE RAET1E SECUENCIADA	49
	6.1 Identificación De Sitios Putativos De Unión A Factores De Transcripción	49
	6.2 Evaluación Del Efecto De Los SNVs En Los Sitios Putativos De Unión A Factores De Transcripción.....	50
7.	RECURSOS MATERIALES Y FINANCIAMIENTO	51
	RESULTADOS	52
	OBJETIVO ESPECÍFICO 1: DETERMINAR LA PRESENCIA DE VARIANTES EN EL PROMOTOR DEL GEN RAET1E EN PACIENTES CON CÁNCER GÁSTRICO.	52
	1. PACIENTES.....	52
	2. EVALUACIÓN CUANTITATIVA Y CUALITATIVA DEL ADN GENÓMICO	54
	3. ANÁLISIS DE LA SECUENCIACIÓN DEL PROMOTOR DE RAET1E	57
	4. IDENTIFICACIÓN DE VARIANTES EN EL PROMOTOR DE RAET1E	59
	OBJETIVO ESPECÍFICO 2: EVALUAR, A TRAVÉS DE ESTUDIOS BIOINFORMÁTICOS, AQUELLAS VARIANTES QUE GENEREN CAMBIOS RELEVANTES EN EL PROMOTOR DE RAET1E.....	65
	1. ELEMENTOS REGULADORES DE TIPO BASAL PRESENTES EN EL PROMOTOR DE RAET1E.	65
	2. ANÁLISIS <i>IN SILICO</i> DE SITIOS PUTATIVOS DE UNIÓN PARA FACTORES DE TRANSCRIPCIÓN (FT).....	65
	3. EFECTO DE LOS SNV _s EN LOS SITIOS DE UNIÓN A FACTORES DE TRANSCRIPCIÓN.....	72
	DISCUSIÓN.....	75
	CONCLUSIONES.....	83
	ANEXOS.....	84
	ANEXO A. FRECUENCIAS ALÉLICAS DE LAS VARIANTES: SNV3, SNV11, SNV42 Y SNV32.....	85
	ANEXO B. ACTA APROBACIÓN Y CONSENTIMIENTO INFORMADO PACIENTES DIAGNOSTICADOS CON CG, DEL COMITÉ ÉTICA EN INVESTIGACIÓN EN SERES HUMANOS, FACULTAD MEDICINA, UNIVERSIDAD DE CHILE.	88
	ANEXO C. RESUMEN CONGRESO GENERADO POR LA TESIS.....	92
	BIBLIOGRAFÍA.....	93

ÍNDICE DE FIGURAS

Figura 1. Mortalidad por cáncer gástrico a nivel mundial.....	16
Figura 2. Localización de los genes que codifican para los ligandos de NKG2D en humanos	20
Figura 3. Dominios proteicos y estructura de los ligandos de NKG2D.....	21
Figura 4. Panorama de los mecanismos de regulación transcripcional descritos para los ligandos de NKG2D.....	24
Figura 5. Esquema de los promotores del gen <i>RAET1E</i>	33
Figura 6. Predicción de la estructura de la cromatina para los promotores descritos del gen <i>RAET1E</i>	33
Figura 7. Representación gráfica de la preparación de librerías genéticas mediante el “Truseq Custom Amplicon”	43
Figura 8. Representación de los archivos FASTQ generados durante la secuenciación en el equipo MiSeq	45
Figura 9. Flujo de trabajo empleado para el análisis de bioinformático de los datos obtenidos en la secuenciación del promotor de <i>RAET1E</i>	47
Figura 10. Archivo VCF final con las variantes que pasaron todos los filtros identificadas para el promotor de <i>RAET1E</i>	49
Figura 11. Evaluación Cuantitativa y Cualitativa del ADN genómico extraído de tumores de pacientes con adenocarcinoma gástrico.....	55
Figura 12. Flujo de trabajo empleado en la evaluación cualitativa y cuantitativa del ADNg de los pacientes en estudio.	56
Figura 13. Esquema de los amplicones que mapean la región de estudio y análisis de su cobertura promedio.	58
Figura 14. SNVs identificados en el promotor de <i>RAET1E</i> y su distribución en la región estudiada.	59
Figura 15. Predicción de sitios de unión putativos para factores de transcripción en la posición de los SNVs identificados a partir de la secuencia no mutada del promotor de <i>RAET1E_1</i>	71
Figura 16.Efecto de los SNVs en los sitios putativos de unión de factores de transcripción.....	74

ÍNDICE DE TABLAS

Tabla 1. SNPs en alelos de MICA asociados a enfermedades.....	29
Tabla 2. Isoformas de RAET1E en humanos.	31
Tabla 3. Protocolo preparación curva estándar para ensayo PicoGreen.	39
Tabla 4. Filtro aplicados para la selección de variantes en el promotor de RAET1E.	48
Tabla 5. Características Clínico-patológicas de los pacientes con Cáncer Gástrico.	53
Tabla 6. Distribución de los casos según Invasión Ganglionar.	53
Tabla 7. Distribución de los casos según Metástasis a distancia.	53
Tabla 8. SNVs identificados en el promotor de RAET1E (RAET1E_1) en pacientes con adenocarcinoma gástrico.	60
Tabla 9. Frecuencias Alélicas de las variantes germinales identificadas en el promotor de RAET1E.	63
Tabla 10. Frecuencias Alélicas de las variantes somáticas identificadas en el promotor de RAET1E.	64
Tabla 11. Sitios putativos de unión a FTs, identificados en la hebra positiva de la región promotora de RAET1E (Predicción realizada mediante PROMO 3.0).	66
Tabla 12. Sitios putativos de unión a FTs, identificados en la hebra negativa de la región promotora estudiada de RAET1E, (Predicción realizada mediante PROMO 3.0).	68
Tabla 13. Sitios putativos de unión a FTs, identificados en la región promotora estudiada de RAET1E, (Predicción realizada con MatInspector).	70

ABREVIATURAS

CG: Cáncer Gástrico

ASR: Age-standardized rate (Tasa de mortalidad)

GIST: GastroIntestinal Stromal Tumor

NK: Natural Killer

MHC: Mayor Histocompatibility Complex

NKG2D: Natural Killer Group 2 member D

NKG2DL: Natural Killer Group 2 member D Ligands

MICA/B: MHC class I polypeptide-related sequence A/B

RAET1: Retinoic acid early Transcription 1/Transcritos tempranos inducidos por ácido retinoico.

RAET1E: Retinoic Acid Early Transcript 1E

ULBP: UL16-binding protein / Proteínas de unión a UL-16

GPI: Glicosil Fosfatidil Inositol

TM: Transmembrana/Transmembranal

TLR: Toll like Receptor / Receptor tipo Toll

Pol2: Polimerasa 2

CAT: Custom Amplicon Oligo Tube (Illumina)

ACD1: Amplicon Control DNA 1

ACP1: Amplicon Control Oligo Pool 1

INDELS: Inserciones y Deleciones.

VCF: Variant Caller Files.

FT: Factor de Transcripción

ADNg: ADN genómico

RESUMEN

En Chile el cáncer gástrico es la primera causa de muerte por tumores malignos para ambos sexos, con alrededor de 3000 personas fallecidas al año, describiéndose las mayores tasas de mortalidad entre la VII y X región y las menores en las regiones extremas.

Diversas investigaciones en el campo de la inmunidad anti-tumoral, han permitido evidenciar el rol del sistema inmune tanto en el control de células tumorales en etapas tempranas como en la evasión de la destrucción inmune favoreciendo el desarrollo del cáncer. Dentro de las células inmunes, las células *Natural Killer* (NK) tienen la capacidad de reconocer y eliminar células tumorales sin haber sido sensibilizadas previamente con el antígeno. Esto ocurre vía su receptor de activación NKG2D, cuya activación dependerá de la unión de sus ligandos expresados en la superficie de células tumorales. El reconocimiento y la unión del receptor NKG2D a sus ligandos (NKG2DLs) permite la activación de las células NK, conduciendo a la liberación de citoquinas o gránulos citolíticos y con ello la muerte de la célula tumoral que expresa uno o varios NKG2DLs en su membrana. Sin embargo, la generación de altos niveles de NKG2DLs en la superficie celular, así como la liberación de la molécula en forma soluble al ambiente tumoral, regulan negativamente la activación de las células NK afectando su funcionalidad, ya que inducen una disminución de NKG2D en la superficie de las células citolíticas, permitiendo con ello el crecimiento tumoral. Adicionalmente, los ligandos de NKG2D liberados compiten con las moléculas en la superficie de células tumorales, alejando a las células citolíticas de sus células blanco.

En humanos se han descrito 2 familias de ligandos de NKG2D: las moléculas MIC y RAET1. Varios polimorfismos han sido descritos para estos ligandos, algunos de los cuales han sido asociados con un aumento del riesgo a desarrollar cáncer y podrían estar implicados como potenciales mecanismos de evasión inmune. Principalmente aquellos que generan cambios en residuos que disminuyen la afinidad por NKG2D o facilitan la liberación del ectodominio del NKG2DL.

El gen *RAET1E* se caracteriza por presentar una identidad de secuencia menor al 43% con el resto de la familia de ligandos de NKG2D en humanos, lo que podría deberse a un rol distintivo de esta molécula. Su expresión en algunos tipos de cáncer se ha identificado como un predictor de mal pronóstico. Además, esta molécula se expresa en tumores de pacientes y líneas celulares de cáncer gástrico, mostrando un patrón de expresión heterogéneo. Llamativamente, estudios previos mostraron que la expresión de *RAET1E* en líneas celulares de CG, frente al estímulo de *Helicobacter pylori*, no fue revertida al bloquear MyD88, molécula corriente abajo de los TLR a diferencia de lo que ocurrió con los otros ligandos de NKG2D. Lo que podría deberse a que exista otra vía que regule su expresión.

RAET1E, es el gen más variable dentro de los otros ligandos de la familia *RAET1*. La existencia de variaciones genéticas en regiones regulatorias que alteren la transcripción ha sido establecida durante mucho tiempo y tendrían un rol en las enfermedades humanas, entre ellas el cáncer. Estas variaciones podrían alterar sitios de unión a factores de transcripción y por tanto modificar su expresión, en particular, en condiciones de estrés como ocurre en un proceso tumorigénico. Se han descrito dos promotores para el gen *RAET1E*, siendo el 1 el más activo transcripcionalmente.

Por lo tanto nos hemos propuesto como hipótesis, que “variantes en el promotor 1 de *RAET1E* en tumores de pacientes con adenocarcinoma gástrico modifican los sitios de unión para factores de transcripción”. Entonces, el objetivo general de este estudio fue “identificar variaciones en la región promotora de *RAET1E* que generaren cambios en esa región reguladora del gen, específicamente, alterando sitios de unión a factores de transcripción en pacientes con cáncer gástrico. Para ello, se analizó la región promotora de *RAET1E* en 50 muestras de adenocarcinoma gástrico a través de secuenciación de última generación. Y además, se evaluó *in silico* el efecto de estas variaciones.

Identificamos 42 variantes de nucleótido único (SNV) en los tumores de pacientes con cáncer gástrico, 19 de ellas no reportadas previamente. No todos los SNVs generaron la modificación de sitios de unión para factores de transcripción

(FT). Se determinó que 9 SNV alterarían sitios de unión para factores de transcripción. Entre los sitios afectados se generarían nuevos sitios de unión para los FTs: NKX 2-5, p53, AP-1, EVI1 y NFkB, mientras que se suprimirían sitios de unión para los FTs: MZF1, KLF3, c-Rel, SALL2, WT1 e IRF-2. Algunas de estas alteraciones de encontrarían presentes simultáneamente en un mismo paciente.

En consecuencia, creemos que la alteración de estos sitios de unión podría generar cambios en la regulación del gen *RAET1E* al ser inducido por algunos de estos factores de transcripción, alterando la respuesta inmune mediada por este ligando y favoreciendo mecanismos de evasión en los pacientes con adenocarcinoma gástrico.

ABSTRACT

In Chile, gastric cancer (GC) is the leading cause of death due to malignant tumors in both sexes, with near 3,000 fatalities each year. The highest mortality rates have been described between VII and X regions while the lower rates are in the extreme regions.

Several years of research in anti-tumour immunity have evidenced the role of the immune system, both in controlling tumor cells in early stages as in the evasion of immune-mediated destruction, promoting cancer development. Natural Killer (NK) cells have the ability to recognize and eliminate tumor cells without having previously been sensitized with the antigen. This occurs through the NKG2D activating receptor, whose activation depends on the union of their ligands expressed on the surface of tumor cells. The recognition and binding receptor NKG2D to its ligands (NKG2DLs) allows the activation of NK cells, leading to cytokine release or cytolytic granules and thereby death of the tumor cell expressing one or more NKG2DLs in their membrane. However, the expression of high levels of NKG2DLs on the cell surface and release of the molecule as soluble tumor environment, downregulate activation of NK cells affecting their functionality, since they induce a decrease NKG2D in cytolytic cell surface, thereby allowing tumor growth. Additionally, soluble NKG2D ligands compete with molecules on the surface of tumor cells, killer cells away to their target cells.

2 families of human NKG2D ligands have been described: the MIC and RAET1 molecules. Several polymorphisms have been described for these ligands, some of which have been associated with an increased risk to develop cancer and could be implicated as potential immune evasion mechanisms. Mainly those that generate changes that reduce waste affinity for NKG2D or facilitate the release of the ectodomain of NKG2DL.

The *RAET1E* gene is characterized as the most divergent within the family of NKG2D ligands in humans, which could be due to a distinctive role of this molecule. Expression in some types of cancer has been identified as a predictor of poor prognosis. Furthermore, this molecule is expressed in tumors of patients and gastric cancer cell lines, showing a pattern of heterogeneous expression. Interestingly,

previous studies showed that expression of RAET1E cell lines CG, in response to stimulus with *Helicobacter pylori* was not reversed by blocking MyD88, a molecule downstream of the TLR, unlike what happened with the other ligands of NKG2D. This could be explained by another pathway regulating its expression.

The *RAET1E* gene is the most polymorphic among RAET1 ligands. The existence of genetic variations in regulatory regions that alter transcription has been established for a long time and has been shown to have a role in human diseases, including cancer. These variations could alter binding sites of transcription factors and thus alter its transcription and expression, in particular under stress as occurs in a tumorigenic process. Two promoters have been described for RAET1E gene and the first promoter (RAET1E_1) has been predicted as the most active transcriptionally.

RAET1E_1 promoter variants in tumors of patients with gastric adenocarcinoma modify binding sites for transcription factors. The aim of this study was to identify variations in the promoter RAET1_1 that generate changes in its regulatory region and that may explain evasion of the immune response in tumors of patients with gastric cancer. Specifically, altering binding sites of transcription factors in patients with gastric cancer. For this purpose, the *RAET1E* promoter region was analyzed in fifty samples from patient with gastric adenocarcinoma through next generation sequencing. In addition, we analyzed the effect of these variations by computational approaches.

We identified 42 single nucleotide variants (SNV) from tumors of patients with gastric cancer. Nineteen SNVs have not previously been reported. Only nine SNVs could alter binding sites for transcription factors. Transcription factors binding sites generated were: NKX 2-5, p53, AP-1, NFkB and EVI1 while transcription factor binding sites deleted were: MZF1, KLF3, c-Rel, SALL2, WT1 and IRF-2. Some of these alterations were simultaneously present in a same patient.

Therefore, we believe that the alteration of these binding sites could lead to changes in the regulation of *RAET1E* gene when induced by some of these transcription factors, altering the immune response mediated by the ligand and favoring evasion mechanisms in patients with gastric adenocarcinoma.

INTRODUCCIÓN

1. CÁNCER GÁSTRICO

De acuerdo a los últimos informes de cifras mundiales de cáncer, esta enfermedad multifactorial sigue encabezando el listado como una de las principales causas de morbilidad y mortalidad a nivel mundial. El 2012 hubo unos 14 millones de nuevos casos y 8,2 millones de muertes relacionadas con el cáncer, y se prevé que esta cifra siga en aumento en los próximos 20 años¹.

Cáncer de pulmón, de mama, de colon, de próstata y de estómago, figuran como los 5 tipos de cáncer más frecuentes en el mundo. En particular, el cáncer de estómago o gástrico (CG) corresponde a nivel mundial al cuarto tipo de cáncer más frecuente en hombres y al quinto en mujeres². Sin embargo, se sitúa como la segunda causa de muerte por cáncer en el mundo con aproximadamente 738.000 muertes/año³.

En relación a la distribución geográfica del CG, resulta llamativa la variabilidad que presenta este tipo de cáncer en los distintos países y regiones del planeta. La mayor incidencia de CG se presenta en Asia, América Latina y el Caribe, mientras que los índices más bajos se concentran en África y Norteamérica. En particular, los países con mayores índices de mortalidad por CG en Asia corresponden a Mongolia seguido de Kyrgiztan y Tajikistan. Mientras que en América Latina y el Caribe los países con las más altas tasas de mortalidad son Guatemala, Ecuador, Chile y Perú⁴ (*Figura 1*).

En Chile el cáncer gástrico es la primera causa de muerte por tumores malignos para ambos sexos, con alrededor de 3000 personas fallecidas al año, describiéndose las mayores tasas de mortalidad entre la VII y X región y las menores en las regiones extremas⁵.

Se han identificado varios factores de riesgo para el CG, entre ellos factores sociodemográficos tales como la edad y el género. Las estadísticas evidencian que el CG es tres veces más frecuente en hombres que en mujeres y aun cuando puede presentarse a cualquier edad, su frecuencia bajo los 40 años es muy baja alcanzando su mayor incidencia entre los 60 y 80 años⁶. Hábitos alimenticios como el alto consumo de productos ahumados, salados, nitrados y carbohidratos, así como la baja ingesta de frutas y verduras, aumentarían significativamente el riesgo de desarrollar CG. Por otro lado, un factor que ha adquirido mayor relevancia los últimos años corresponde a la infección por *Helicobacter pylori*. Varios estudios epidemiológicos han proporcionado evidencia de la asociación entre la infección por *H. pylori* y el desarrollo de CG⁷.

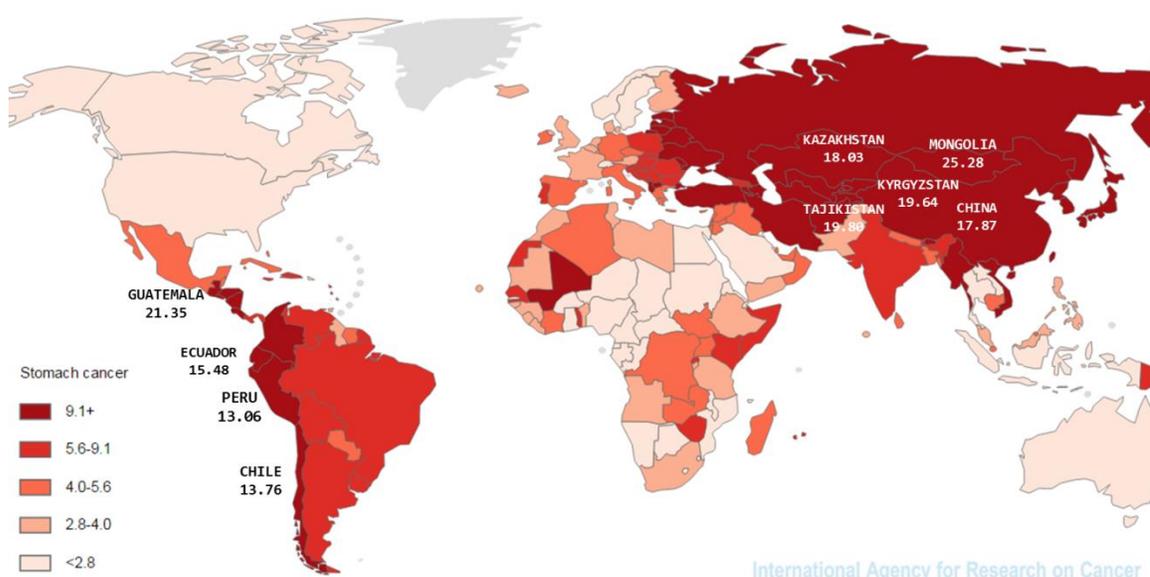


Figura 1. Mortalidad por cáncer gástrico a nivel mundial. El mapa muestra la distribución y tasa mortalidad en las diferentes regiones del mundo incluyendo a ambos sexos. El indicador empleado es el ASR (del inglés “Age-standardized rate”) y corresponde a la tasa de mortalidad por cada 100000 habitantes por año. Sólo se muestran aquellos ASR de los países con mayores índices de mortalidad. Tomado de Ferlay y Cols., 2013,⁴.

De los tipos de cáncer gástrico el 90% de los casos corresponde a adenocarcinoma gástrico, el 10% restante se categoriza como linfomas gástricos, tumores del estroma gastrointestinal (GIST) o sarcomas⁸. Los adenocarcinomas pueden ser de dos tipos: intestinal o difuso. En el primer caso las células tumorales forman glándulas funcionales que se parecen a la mucosa intestinal, mientras que el segundo tipo, las células tumorales pierden la organización, y es más propenso a metástasis, por lo que se asocia a un mal pronóstico⁹.

Dado el carácter asintomático de este cáncer, la mayoría de los pacientes se diagnostican en etapas avanzadas, cuando la resección curativa es prácticamente imposible, lo cual también conduce a un mal pronóstico. El método más sensible para la detección de CG es el *screening* a través de una endoscopia digestiva alta. Las opciones de tratamientos para el 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 linfadenectomía¹⁰. Normalmente, el adenocarcinoma gástrico progresa de manera lenta y asintomática y en Chile más de la mitad de los pacientes se encuentran en un estado avanzado de la enfermedad al momento del diagnóstico⁵. Además, entre el 40 y 65% de los pacientes resecados con intento curativo presentarán recurrencia de la enfermedad. Debido a lo anterior, la sobrevivida a 5 años de los pacientes con CG resecable varía entre un 10 y 30% con un tiempo de sobrevivida promedio de 24 meses⁹. De esta forma, el cáncer gástrico es un tema prioritario en cuanto a salud pública en nuestro país.

2. INMUNIDAD ANTITUMORAL Y NKG2D

Varios años de investigación, en el campo de la inmunidad anti-tumoral, han permitido evidenciar el rol del sistema inmune tanto en el control de células tumorales en etapas tempranas como en la evasión de la destrucción inmune favoreciendo el desarrollo del cáncer¹¹.

Un grupo particular de células con la capacidad de reconocer y eliminar células tumorales sin haber sido sensibilizadas previamente, corresponde a las células *Natural Killer (NK)*¹². Estas células forman parte del sistema inmune innato y proporcionan protección tanto contra patógenos como tumores. Su función está fuertemente regulada por señales inhibitorias y activadoras que simultáneamente le están siendo proporcionadas en el microambiente celular¹³. El balance de estas señales captadas y procesadas por receptores distribuidos en la superficie celular promoverán o no su actividad citolítica. Para su activación la célula NK reconocerá a través de receptores activadores, moléculas propias en la superficie de las células blanco cuya expresión se induce en condiciones de estrés, como la generada durante un proceso infeccioso o tumorigénico¹⁴, en cuyo caso se activará la respuesta citotóxica y secretora de citoquinas. Mientras que el estado de tolerancia ocurre tras la unión de moléculas del complejo MHC de clase I a sus receptores de inhibición.

Dentro de los receptores de activación identificados en la células NK, NKG2D (*del inglés, Natural Killer group 2, member D*) ha sido ampliamente estudiado por su rol en la inmunovigilancia anti-tumoral. La expresión de este receptor no es exclusiva de células NK, también se ha identificado en otros linfocitos como las células NKT (*del inglés Natural Killer T Cells*), T citotóxicos CD8+, T $\gamma\delta$, y bajo ciertas condiciones por células T CD4+. Experimentos *in vivo*, demuestran que la sobreexpresión de los ligandos de este receptor en células cancerosas causa el rechazo de tumores en ratones¹⁵. Así mismo, estudios empleando ratones deficientes en la molécula NKG2D, muestran que en comparación con los ratones *wild type*, la deficiencia del receptor se relaciona con un incremento en la tasa de desarrollo tumoral¹⁶. En relación al CG, estudios desarrollados por Saito y Cols.¹⁷ muestran

que la disfunción de las células NK, en pacientes con CG, se correlacionaría con la disminución de la expresión del receptor de activación NKG2D. A su vez, otro estudio mostró también una disminución en la expresión de NKG2D en linfocitos T CD8+ circulantes, siendo este uno de los mecanismos de evasión del sistema inmune por parte del tumor en este tipo de cáncer^{17,18}. Por lo tanto, el receptor NKG2D y sus ligandos son moléculas claves para el control tumoral.

3. LIGANDOS DEL RECEPTOR DE ACTIVACIÓN NKG2D

A la fecha, múltiples ligandos del receptor NKG2D (NKG2DL) han sido identificados tanto en humanos como en ratón. Todos son homólogos de la molécula del complejo mayor de histocompatibilidad de clase I, pero a diferencia de esta, no se asocian con β_2 -microglobulina, ni tienen algún rol en la presentación antigénica.

De acuerdo a su organización genética, en humanos estos ligandos han sido agrupados en dos grandes familias: los genes MIC (*del inglés MHC class I-related genes*), y los genes RAET1 (*del inglés Retinoic acid early Transcription 1*) también conocidos como *ULBP* (*del inglés UL16-binding protein*), ambas familias de genes se localizan en el cromosoma 6¹⁹ (*Figura 2*).

Los genes *MIC* están localizados dentro del complejo principal de histocompatibilidad mayor en el brazo corto en la región 6p21.3 (*Figura 2A*) y se componen de 7 miembros (*MICA*, *MICB*, *MICC*, *MICD*, *MICE*, *MICF* y *MICG*), de los cuales solo *MICA* y *MICB* producen transcritos funcionales (*Figura 2B*). Ambos genes son inusualmente largos (8799 pb. *MICA* y 11716 pb *MICB*) comparado con el tamaño promedio de los genes *HLA-A* a *HLA-G* de (3500 pb). Por otro lado, los genes *RAET1* (*ULBP*) se localizan en el brazo largo en la región comprendida entre las bandas 6q24.2 y 6q25.3. Allí un *cluster* genómico de 180 kb contiene 10 miembros de la familia *RAET1* (*RAET1E*, *RAET1F*, *RAET1G*, *RAET1H*, *RAET1I*, *RAET1J*, *RAET1K*, *RAET1L*, *RAET1M* y *RAET1N*), de los cuales sólo 6 codifican proteínas funcionales. El tamaño de estos genes varía entre los 2407 y 6076 pb.²⁰ (*Figura 2B*).

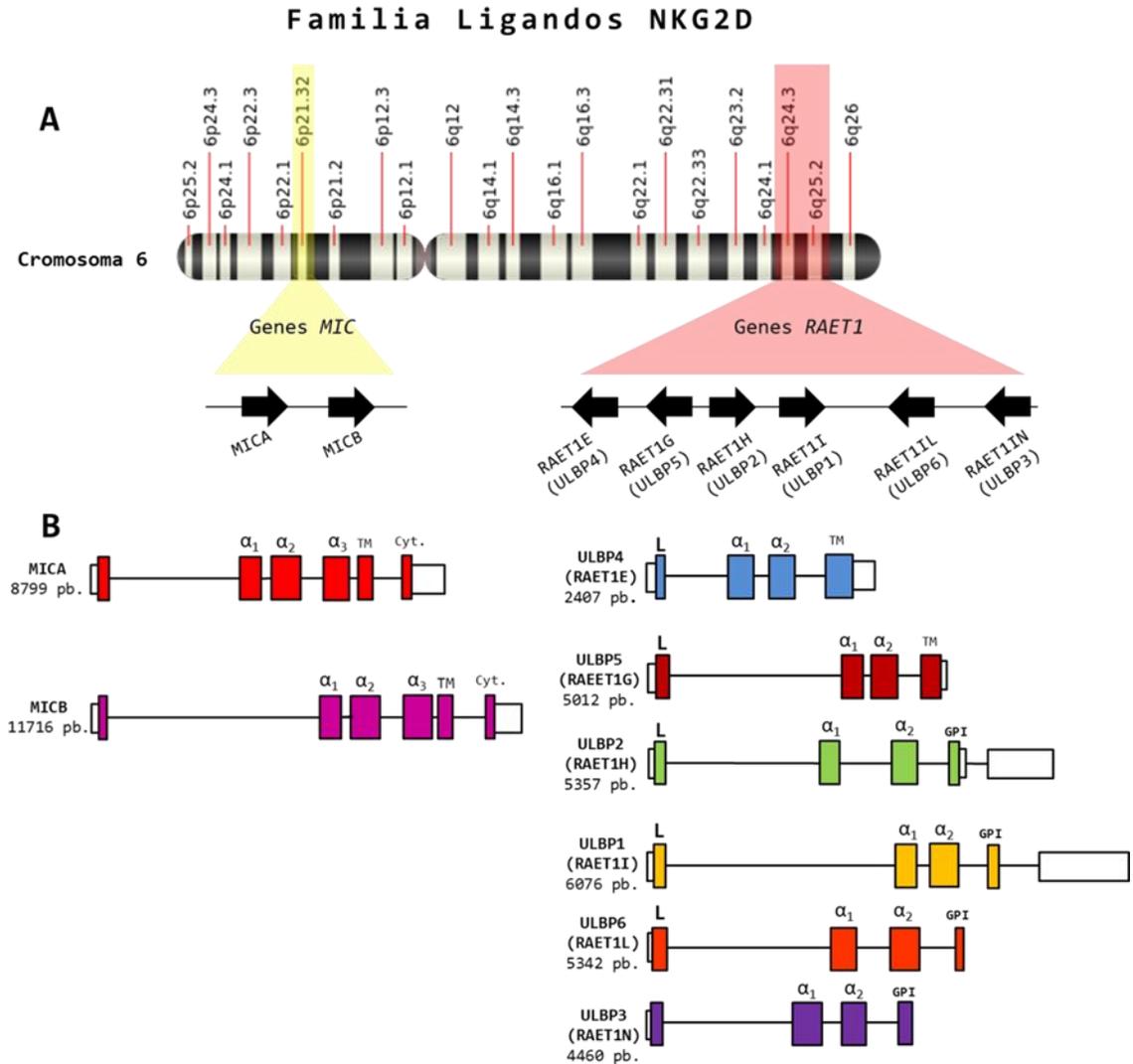


Figura 2. Localización de los genes que codifican para los ligandos de NKG2D en humanos. (A) La familia de genes *MIC* y *RAET1/ULBP* se localizan en el brazo corto (6p21.3) y largo (6q24.2-q25.3) del cromosoma 6, respectivamente. Las flechas negras indican la posición relativa y orientación de los genes. **(B)** Se ha representado la longitud relativa de los genes y la organización de sus exones codificantes (cajas de colores) así como los exones no codificantes (cajas en blanco). Las abreviaciones L, GPI, TM y Cyt corresponden a las secuencias codificantes del péptido líder, el sitio de anclaje glicosilfosfatidilinositol, dominio transmembrana y región citoplasmática, respectivamente.

Aunque dentro de cada familia de NKG2DL son altamente conservados, los miembros de la familia MIC comparten una baja similitud de secuencia o estructura con la familia RAET1^{20,21}. Estructuralmente, los NKG2DL corresponden a glicoproteínas de membrana. Las moléculas MIC consisten de tres dominios extracelulares (α_1 , α_2 y α_3), seguidos de una región transmembrana, y una cola citoplasmática. Por su parte las moléculas RAET1 pierden el dominio extracelular α_3 , y algunas moléculas como RAET1 I, N y L, están ancladas a la membrana a través de un tallo Glicosil-fosfatidil-inositol (GPI), mientras que las proteínas RAET1E y G poseen regiones transmembrana (*Figura 3*). En cuanto a la molécula RAET1H se ha descrito que puede estar anclada por un tallo GPI o bien ser una proteína transmembrana²².

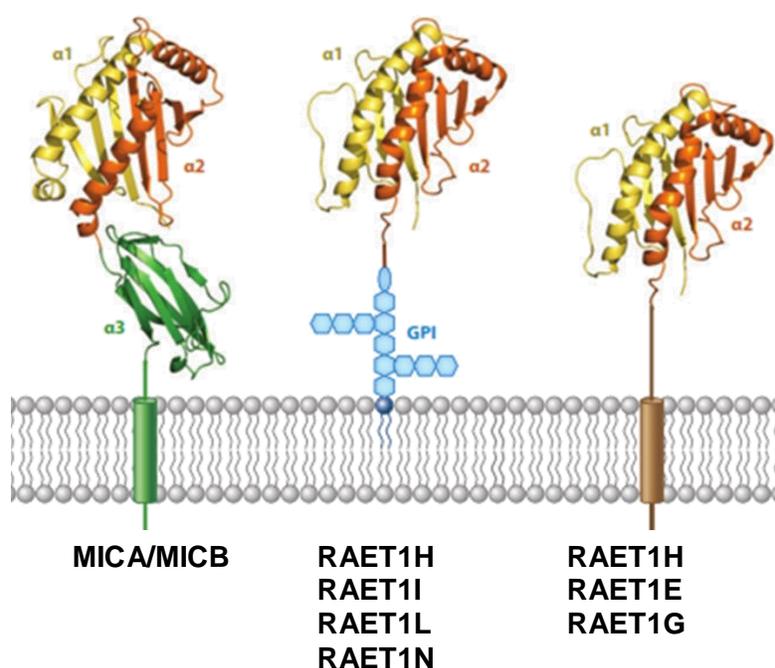


Figura 3. Dominios proteicos y estructura de los ligandos de NKG2D. La figura esquematiza la estructura de los diferentes ligandos descritos a la fecha para el receptor de activación NKG2D. Los dominios relevantes en la interacción ligando-receptor corresponden a los dominios α_1 y α_2 , que según se observa son muy similares estructuralmente entre los diferentes NKG2DL. Destaca la ausencia del dominio α_3 en los genes RAET1, así como variaciones en la vinculación con la membrana. Tomada y modificada de Raulet y col. 2013²¹.

4. EXPRESIÓN DE LOS LIGANDOS DE NKG2D

En condiciones normales la expresión de los NKG2DLs en las células es baja o nula. Sin embargo, en la literatura se han reportado casos donde existe expresión pese a haber ausencia de estrés celular. Por ejemplo, en humanos, RAET1I (ULBP1) ha sido detectado en varias células sanguíneas como linfocitos B, plaquetas, monocitos y granulocitos, por lo cual se cree que esta molécula podría participar en la hematopoyesis. En otro estudio se reportó que los ligandos: MICA, RAET1I y RAET1N (ULBP3) en humanos, así como los ligandos: Rae-1 y H60 en ratones, se expresan en las células de la médula ósea incluyendo las células estromales. La molécula RAET1N sido reportada por su expresión en células madres mesenquimales humanas^{20,23,24} y las moléculas MICA y MICB han sido identificados por su expresión en el epitelio intestinal, probablemente debido a la estimulación de la flora bacteriana en el intestino²⁰. Es importante destacar que en muchos de estos estudios se reporta la expresión en base a la cuantificación del ARNm, que no siempre se correlaciona con la presencia de la proteína en la superficie celular.

Una amplia variedad de tumores y líneas celulares de tumor expresan los ligandos de NKG2D. Al respecto Carapito y Bahram²⁰ señalan que esta expresión sería adquirida en el proceso tumorigénico y no durante la transformación, porque la expresión de oncogenes clásicos como K-ras, c-myc o Akt en células de ovario murinas no produjeron un aumento en los niveles de los ligandos²⁵. Así mismo, cada tumor en particular presenta un patrón específico de expresión de los ligandos. Las moléculas MICs se encuentran más frecuentemente en tumores epiteliales derivados de tejidos como ovarios o colon y en menor medida en neoplasias hematológicas. En contraste, la familia RAET1 ha sido identificada mayoritariamente en líneas celulares de leucemia de linfocitos T, y detectada en células leucémicas aisladas de pacientes²⁰.

En nuestro laboratorio hemos analizado la expresión de los NKG2DL tanto en muestras tumorales de pacientes con cáncer gástrico como en líneas celulares de adenocarcinoma gástrico. Los ensayos de inmunohistoquímica muestran que existe una expresión significativamente mayor de MICA y MICB en células tumorales que

en mucosa adyacente y en ambas mayores que en muestras de mucosa gástrica de pacientes sometidos a cirugía bariátrica, utilizados como control no tumoral. Por otro lado, los análisis de expresión de estas moléculas, mediante ensayos de citometría de flujo, muestran que los NKG2DL, entre ellos RAET1E, se expresan tanto en células derivadas de adenocarcinoma gástrico como en la mucosa gástrica adyacente al tumor. Los análisis muestran que en el 85% de los casos analizados, sobre el 50% de las células expresan uno o todos los ligandos estudiados. No se observaron diferencias estadísticamente significativas en la expresión entre el tumor y la mucosa gástrica. Esto podría deberse a que entre los pacientes hay una expresión muy heterogénea de los distintos ligandos. Sólo un 40% de las muestras expresan estadísticamente más NKG2DL en células tumorales que en mucosa adyacente [Karina Kramm, Tesis para optar al título de Magister en Ciencias].

Por último, una expresión aberrante de los ligandos ha sido vinculada a enfermedades autoinmunes e inflamatorias como artritis reumatoide o la enfermedad de Crohn²⁰.

5. REGULACIÓN DE LA EXPRESIÓN DE LOS LIGANDOS DE NKG2D.

En relación a la regulación de estos ligandos se sabe que ésta ocurriría en varias etapas de su biogénesis y que las vías de respuesta al estrés serían las responsables de la inducción de su expresión, tanto en condiciones fisiológicas (nutricionales/físicas) como patológicas (oncogénicas/ infección microbiana)^{21,26}. Vías de estrés activadas por infecciones, choque térmico, procesos tumorigénicos, daño tisular y daño genotóxico regularían la transcripción, traducción y estabilización del ARNm o proteína, y la expresión en superficie de los ligandos de NKG2D. Dado el interés de este estudio nos enfocaremos en los mecanismos de regulación descritos a nivel transcripcional para los ligandos de NKG2D (figura 4).

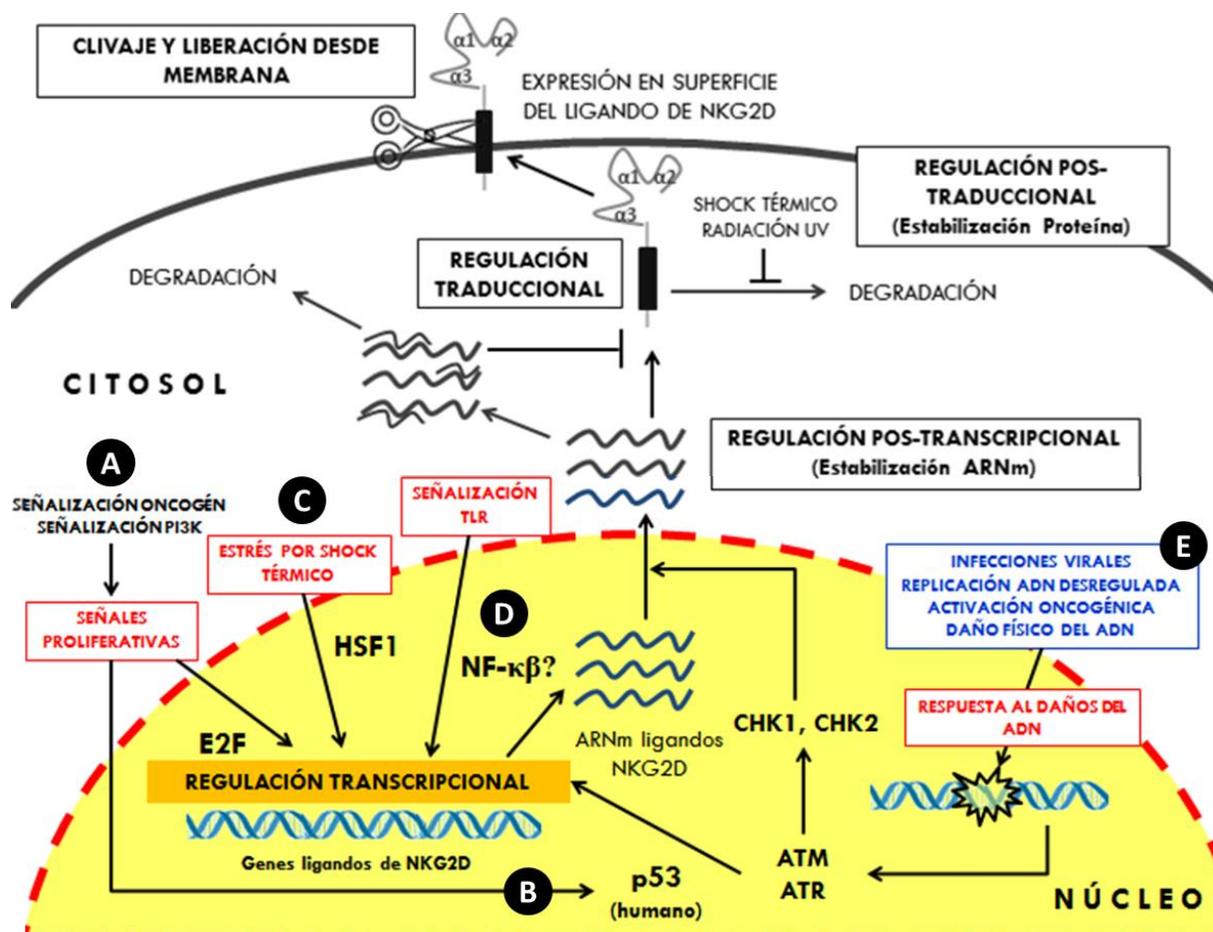


Figura 4. Panorama de los mecanismos de regulación transcripcional descritos para los ligandos de NKG2D. Como vemos la regulación de la expresión de los ligandos ocurre en varias etapas de su biogénesis. A nivel transcripcional varias son las señales que estaría induciendo su expresión: **(A)** Señales proliferativas, que pueden ser activadas por oncoproteínas o la señalización de PI3K, pueden inducir la transcripción de los genes *Rae-1*, *RAET1* y *MICA*. En el caso de *Rae-1* es mediado por la activación y/o inducción de los factores de transcripción E2F. **(B)** La activación oncogénica puede indirectamente activar a p53 aumentando la transcripción de algunos ligandos de NKG2D como *RAET1I* y *RAET1H*. **(C)** El estrés por shock térmico resulta en la activación de los promotores de *MICA* y *MICB*, a través del factor de transcripción HSF1. **(D)** NF-κβ, también se ha reportado que se une y transactiva a los genes de los NKG2DL, aumentando la posibilidad de que la expresión inducida por TLRs sea mediada por la activación de NF-κβ. A nivel pos-transcripcional operan otras vías. **(E)** La respuesta al daño del ADN, inducida por el daño al ADN o una replicación desregulada, induce la expresión de los ligandos, probablemente regulando factores que estabilizan el ARNm. En este nivel la inducción de los ligandos es dependiente de las quinasas que censan el daño del ADN como ATM, así como las quinasas corriente abajo CHK1 y CHK2. Tomado y modificado de Iannello y col., 2013²⁷.

5.1 Regulación Transcripcional

Los promotores de los genes que codifican MICA y MICB poseen elementos de respuesta al choque térmico, por ello se ha descrito que la expresión de ambos ligandos es regulada por esta vía y, a través de ensayos de inmunoprecipitación de cromatina (ChIP), se mostró que tanto el promotor del gen de *MICA* como *MICB* fueron ocupados por el factor de transcripción de choque térmico 1 (HSF1)²⁸.

Por otro lado, Iannello y Raulet²⁹ demostraron que las condiciones proliferativas *in vitro* podrían inducir la transcripción de algunos ligandos del receptor NKG2D, incluyendo a la familia de ligandos Rae-1 en fibroblastos murinos proliferantes, y de MICA, MICB, RAET1H y RAET1N en la línea celular humana HCT116²¹. La inducción del gen *Rae-1* en células proliferantes se debe a la familia de factores de transcripción E2F, claves reguladores de la entrada al ciclo celular²⁹.

Dado que p53 es un potente supresor de tumores, numerosos estudios han investigado su rol en la regulación de los NKG2DL. Los primeros trabajos reportaron que la expresión de estas moléculas se observaba en muchas líneas celulares, incluso cuando el gen p53 era suprimido, sugiriendo entonces que p53 no era esencial para la expresión de los ligandos²⁵. Sin embargo, un análisis de los ligandos RAET1I y RAET1H sugiere que p53 podría potenciar la expresión de estos ligandos. En este sentido elementos de respuesta a p53 se localizaron en los primeros intrones de estos genes, y en una línea celular se observó, mediante un ensayo ChIP, que p53 se asocia a los genes *RAET1I* y *RAET1H*, por lo cual se cree que la acción de p53 podría amplificar la transcripción de algunos NKG2DL^{21,26}. Contrario a este efecto, otro reporte indicó que microARNs inducidos por p53 pueden regular negativamente a RAET1H³⁰. Entonces, pese a que p53 sea activado corriente abajo del daño del ADN, y aunque la respuesta al daño del ADN induce a los ligandos en células murinas, no puede afirmarse que este supresor de tumores sea relevante en la inducción de la expresión de los NKG2DL. Al respecto, se cree que la inducción de los ligandos en respuesta al daño del ADN sería en parte, por un proceso independiente de p53²¹.

Los factores de transcripción de la familia Sp, así como NF- κ B, han sido reportados por unirse al promotor y participar en la activación transcripcional de los NKG2DL humanos. Además los factores de transcripción AP-1 y AP-2a, estarían involucrados en la regulación negativa de los genes *RAET1E* y *RAET1F*³¹. El rol de NF- κ B es relevante dado que sugiere que varios NKG2DL serían regulados por receptores tipo *Toll-like* (TLRs). De hecho, existe evidencia de que varios agonistas de los TLRs indujeron la expresión de transcritos de *RAET1* y de la expresión de la proteína en superficie *Rae-1* en cultivos de macrófagos murinos²¹. Esta inducción de *Rae-1* por ligandos de TLR fue suprimida en ratones carentes de *MYD88*, el principal adaptador de la señalización de los TLRs³².

En el contexto de estos receptores (TLR), en nuestro laboratorio se desarrolló una línea de investigación con el objetivo de contribuir al entendimiento de los mecanismos por los cuales *Helicobacter pylori* regularía la expresión de estos ligandos en CG. Se analizó la expresión de los ligandos de NKG2D en líneas de adenocarcinoma gástrico (AGS, MKN-45 y HTB-135) estimuladas con *Heat killed Helicobacter pylori* (HKHP). Se observó una inducción en la superficie de estas células de las proteínas MICA, y algunos de los ligandos *RAET1* (H, G, L y E).

Dado que el lisado de HKHP contiene diversos patrones moleculares asociados a patógenos (PAMPs) con el potencial de ser reconocidos por los TLRs (TLR4, TLR2 y TLR5), se evaluó el efecto de HKHP sobre la expresión de todos los NKG2DL tras la inhibición de *MyD88*, una molécula adaptadora corriente abajo de la vía de los TLRs. Los resultados mostraron que tras la inhibición hubo una reversión del efecto inductor de la localización en superficie de las moléculas MICA y *RAET1* H, G y L. Sin embargo, en el caso de *RAET1E*, se observó un aumento en la localización en superficie de esta molécula con respecto a las células estimuladas sólo con HKHP. Esto llama la atención, ya que esta inducción podría deberse a una vía alternativa u otro mecanismo aún desconocido [Hernández y col., manuscrito en preparación], por lo cual se requiere mayor investigación.

6. MECANISMOS DE EVASIÓN MEDIADOS POR EL SISTEMA NKG2D Y SUS LIGANDOS.

En base a la evidencia planteada anteriormente, muchos investigadores se han preguntado por el mecanismo por el cual los tumores pueden desarrollarse y persistir si expresan en su superficie NKG2DL. En los últimos años se han descrito varios estudios que demuestran que las células tumorales han desarrollado mecanismos que permiten evadir la inmunidad antitumoral mediada por el sistema NKG2D y sus ligandos. Entre los mecanismos descritos encontramos que existirían tumores capaces de (1) secretar citoquinas supresoras como TGF- β e IL-10^{33,34} que disminuyen la expresión de estas moléculas, (2) secretar metaloproteasas que promoverían la liberación de los NKG2DL en forma soluble^{13,35}, (3) excretar vesículas exocíticas conteniendo uno o más ligandos¹³ y (4) producir microARNs con propiedades inmunosupresoras sobre los ligandos de NKG2D^{30,36}. Algunas de estas estrategias estarían asociadas a la disminución de la expresión de los NKG2DL o a competir con las moléculas en solución con los ligandos en la superficie de las células. Pero también se han descrito tumores que promoverían el aumento de la expresión de los ligandos, provocando la inactivación de la respuesta mediada por NKG2D. Este mecanismo, descrito tanto para la presencia sostenida en el tiempo de NKG2DLs, en el células tumorales, como de las moléculas solubles, consistiría en la (5) estimulación en forma crónica e induciendo la internalización del receptor NKG2D con la consiguiente alteración de la respuesta citotóxica mediada por este receptor.

7. VARIABILIDAD DE LOS NKG2DLs

7.1 MICA Y MICB

Los ligandos de NKG2D son moléculas bastante polimórficas. A la fecha se han descrito 100 alelos para *MICA* y 40 para *MICB*, los cuales difieren entre individuos y poblaciones. Entre las variaciones genéticas identificadas para la familia *MIC* la mayoría corresponde a SNPs, en su mayoría localizados en los exones que codifican para los dominios extracelulares que interaccionan con el receptor NKG2D. Estas variaciones generan en su mayoría cambios no sinónimos en los residuos

proteicos³⁷, por lo que se cree podrían alterar la interacción con este receptor y en consecuencia la respuesta citotóxica mediada por NKG2D y sus ligandos.

Un grupo de alelos, principalmente la variante *MICA*008*, tiene una inserción de una guanina después de dos o cinco repeticiones del triplete GCT en el exón que codifica la región transmembranal, lo que resulta en un codón de stop prematuro y una región transmembrana trunca del ligando MICA. Esta variante alélica, denominada también *MICA A5.1*, tiene un importante rol en liberación de MICA como una molécula soluble en células tumorales y una localización diferencial en la membrana plasmática²⁰.

A nivel mundial el alelo *MICA*008* es el más frecuente. Sin embargo, en la población de indígenas sudamericanos este alelo posee una frecuencia menor al 10%²⁰. En relación a la población Europea, indígenas sudamericanos, de China y Tailandia los alelos más abundantes serían *MICA*001*, *MICA*020*, *MICA*045* y *MICA*052*, respectivamente. Carapito y Bahram²⁰ señalan que la especificidad de los alelos en regiones específicas del planeta significaría que cientos de alelos dan lugar a proteínas especializadas en un determinado microambiente microbiano, aunque no se descarta la influencia de la migración.

A diferencia de MICA, la molécula MICB, parece tener una distribución alélica más homogénea en la población, basados en la genotipificación de un número menor de individuos en comparación a MICA. El alelo *MICB*005* es el más frecuente en todas las poblaciones, seguido de *MICB*002*, *MICB*004*, *MICB*008*, *MICB*014* y *MICB*003*²⁰.

Interesantemente, algunos de los SNPs se han asociado con enfermedades en determinadas poblaciones (*tabla 1*). Llamativamente, un mismo SNP para MICA (*p.Met129Val*) se asoció a diferentes enfermedades en dos poblaciones distintas, otorgando una mayor susceptibilidad a la enfermedad. En el primer caso el SNP evaluado en 116 pacientes franceses mostró asociación con la enfermedad de injerto contra huésped. En el segundo caso el mismo SNP se asoció a Carcinoma Nasofaríngeo (NPC) tras evaluarlo en 130 pacientes con NPC y 180 individuos sanos del noreste de Tailandia. Este SNP genera un cambio aminoacídico de metionina a

valina en la posición 129 de la proteína MICA, localizándose en el dominio α_2 de la proteína. Aunque ambos tienen carácter hidrofóbico, la modificación producto del SNP produce un cambio en la interacción con el receptor, la cual resulta ser más débil y por lo tanto afecta la función citotóxica mediada por el receptor NKG2D^{37,38}.

Tabla 1. SNPs en alelos de MICA asociados a enfermedades.

SNP	ENFERMEDAD	EFEECTO	POBLACIÓN
<i>MICA p.Met129Val</i>	Enfermedad de injerto contra huésped	Susceptibilidad	Francesa
<i>MICA p.Met129Val</i>	Carcinoma Nasofaríngeo	Susceptibilidad	Tunecina
<i>MICA p.Ile213Thr y p.Arg251Gln</i>	Cáncer Cervical	Protección	Tailandia del Norte

Tomada y modificada de Choy y Cols., 2010³⁷.

Algunos estudios se han focalizado en identificar polimorfismos en la región promotora de estos genes, dado que podrían originar una expresión aberrante de dichos genes y alterar su correcta función. Rodríguez-Rodero y Cols.,³⁹ identificaron un nuevo polimorfismo en el promotor de *MICB*, una delección de 2 pb (66AGdel) que tendría un impacto en la expresión de *MICB*. Efectivamente estudios funcionales permitieron comprobar que dicha variación alteraba el sitio de unión al factor de transcripción Sp1, disminuyendo la actividad transcripcional de *MICB*.

7.2 *RAET1s* (ULBPs)

La mayoría de los estudios genéticos disponibles se han centrado en el estudio de las regiones codificantes de los genes *RAET1*, que han sido evaluadas en población sana y han permitido identificar las frecuencias alélicas entre las diferentes poblaciones⁴⁰⁻⁴². El estudio de estos ligandos se ha restringido a población Caucásica, del Noreste de Tailandia y Amerindios (indígenas *Kolla*) mostrando claras diferencias de frecuencias entre las poblaciones, sugiriendo que los polimorfismos podrían aumentar en respuesta a la presión selectiva divergente en los diferentes entornos geográficos²⁰.

Basados en el estudio de los exones que codifican para los ectodominios α_1 y α_2 , se determinó que los ligandos más polimórficos fueron *RAET1E* (ULBP4), *RAET1G* (ULBP5), *RAET1H* (ULBP2) y *RAET1L* (ULBP6)²⁰. Interesantemente nuestro gen en estudio *RAET1E* es el más variable, y altamente polimórfico en dos

especies de monos del viejo mundo^{20,43}. Sin embargo, el impacto de estos polimorfismos en términos de la función o su significancia en la evolución de los primates aún debe ser esclarecido.

En cuanto a la variabilidad de las regiones reguladoras de estos genes, existen algunos reportes donde se ha evaluado la existencia de polimorfismos en la región promotora de los genes *RAET1*. Antoun y Cols.,⁴² evaluaron la frecuencia de algunos *SNPs* previamente descritos en la región promotora de los genes *RAET1*. Este estudio incluyó muestras de individuos de diferentes poblaciones definidas como: Euro-caucásicos (EC), Afro-caribeños (AC) e Indio-asiáticos (IA). Los *SNPs* estudiados fueron *RAET1E* g.175C>G, *RAET1H* g.651A>C, y *RAET1N* g.169C>G. De estos, tan sólo se observaron diferencias significativas en el *SNP* de *RAET1N* (*ULBP3*), cuando se compararon las poblaciones AC vs EC, y AC vs IA⁴². Lo interesante sería evaluar el efecto que dicha variación genética pueda tener dentro de la región promotora, y por ende en la inducción de la expresión de dicha molécula. Por otro lado, Eagle y Cols.,⁴⁴ analizaron las regiones promotoras putativas de los *NKG2DL* comparando 600 pb corriente arriba del codón de inicio (AUG) de los 8 ligandos descritos en humanos. Dentro de sus hallazgos reportaron que *RAET1E*, nuestro gen de interés, resultó ser el más divergente de los genes *RAET1* compartiendo una similitud menor al 8% con el resto de los genes. Algunas de las variantes identificadas para *RAET1E* suprimieron sitios de unión a factores de transcripción, localizadas 113 y 519 pb corriente arriba del sitio de inicio de la transcripción (+1). Los factores de transcripción afectados por estas variaciones fueron NR5A2 (-113T>C) y HMX3/TCF15 (-519G>C)⁴⁴. Sin embargo, se requieren mayores estudios para evaluar el efecto de estos polimorfismos a nivel funcional.

8. RAET1E (ULBP4)

Dentro de la familia *RAET1* el gen más divergente es *RAET1E (ULBP4)* presentando una identidad de secuencia menor al 43% con el resto de los miembros de la familia⁴⁵. Al igual que *RAET1G* la proteína posee una región transmembrana y una cola citoplasmática hidrofílica, similar a las moléculas MIC⁴².

De acuerdo a la base de datos NCBI⁴⁶, se han identificado a la fecha 4 transcritos originados por splicing alternativo. Sus características así como las diferencias entre cada transcrito se describen en la *tabla 2*.

Tabla 2. Isoformas de RAET1E en humanos.

R A E T 1 E		
TRANSCRITO/ID	TAMAÑO (pb)	DESCRIPCIÓN
NM_139165	958	Codifica isoforma más larga . Transcrito principal. <u>Isoforma 1</u> .
NM_001243325	850	Codifica isoforma más corta , pierde segmento proteico interno comparado con isoforma anterior. <u>Isoforma 2</u> .
NM_001243327	1032	Variante difiere en extremo 3' comparado con isoforma 1 . Cambio en el marco de lectura con diferencias en C-ter en comparación a isoforma 1. <u>Isoforma 3</u> . *
NM_001243328	2099	Variante contiene un exón 5' no codificante adicional, y 1 exón terminal 3' alternativo en comparación a isoforma 1 . Cambio en el marco de lectura generándose una isoforma corta con C-ter distinto. <u>Isoforma 4</u> .

* La isoforma 3 es mencionada en la literatura científica como RAET1E2. Fuente: NCBI, 2015.

En el contexto del estudio de los mecanismos de evasión inmune generados por las células tumorales, Cao y Cols.,⁴⁷ describieron en el año 2007 una nueva isoforma de RAET1E, identificada como RAET1E2 (*Isoforma 3, tabla 2*), que se genera producto de un proceso de *splicing* alternativo y cuyo producto es una proteína soluble, que carece del dominio transmembrana. Esta versión modificada de RAET1E fue capaz de regular negativamente la citotoxicidad mediada por la activación del receptor NKG2D⁴⁷.

En contraste a los otros ligandos de NKG2D el gen *RAET1E* muestra una expresión diferencial en piel⁴⁸, pero también se ha descrito su expresión en cáncer de ovario. Interesantemente la expresión de RAET1E en este tipo de cáncer se correlacionó inversamente con la supervivencia⁴⁹, por lo que fue descrito como un nuevo predictor de mal pronóstico.

El gen *RAET1E* se localiza en orientación reversa (hebra negativa) en el brazo largo del cromosoma 6 (6q24.2-q25.3) (*Figura 2A*) y su tamaño relativo es 2407 pb (desde codón inicio a codón de término)²⁰, organizando en 4 exones y 3 intrones (*Figura 2B*). Basados tan sólo en la secuenciación de los exones 2 y 3 del gen *RAET1E*, que codifican los dominios extracelulares que interaccionan con NKG2D, se han identificado 11 alelos. De estos, *RAET1E*006* y *RAET1E*007*, se identificaron únicamente en individuos del Noreste de Tailandia, mientras que el alelo *RAET1E*009* es único entre los indígenas americanos Kolla⁴⁵.

En relación a la región reguladora de este gen, y basados en la información disponible en la base de datos *Eucaryotic Promoter Database* (EPD)⁵⁰ se han descrito 2 promotores para *RAET1E*. El primero (*RAET1E_1*) localizado entre las coordenadas genómicas 150,215, 266 y 150,219,266, y el segundo (*RAET1E_2*) localizado entre las coordenadas 150,210,042 y 150,214,042 (*Figura 5*) según ensamblaje del genoma *GRCh37*. Ambas regiones promotoras han sido validadas por la presencia de elementos basales de un promotor (por ejemplo: la presencia de la caja TATA) y caracterizadas por la presencia modificaciones en la cromatina que son comúnmente encontradas en promotores transcripcionalmente activos. Entre las modificaciones de la cromatina presentes en la región promotora de *RAET1E* encontramos la mono o tri-metilación de la histona 3, en la lisina 4 (*Figura 6A-B*). De acuerdo a la literatura esta modificación es característica de promotores transcripcionalmente activos⁵¹. Otro elemento a considerar son los sitios de ocupancia de la ARN polimerasa II (*Figura 6C*). A partir de estas consideraciones se predijo una mayor actividad en el promotor *RAET1E_1* que en el *RAET1E_2*, lo que además significa una mayor accesibilidad de factores de transcripción a esa región (*Figura 6*).

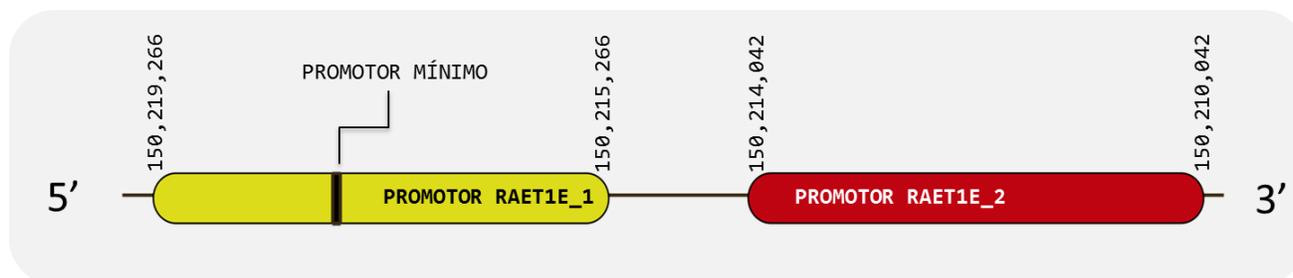


Figura 5. Esquema de los promotores del gen *RAET1E*. La figura muestra la localización de los dos promotores descritos para el gen *RAET1E* (barra amarilla y burdea). En el promotor *RAET1_1* se indica el promotor mínimo, es decir la secuencia que incluye el sitio de inicio de la transcripción o *TSS* (del inglés *Transcription Start Site*). Información colectada de la base de datos EPD, Drees y col. 2014⁵⁰

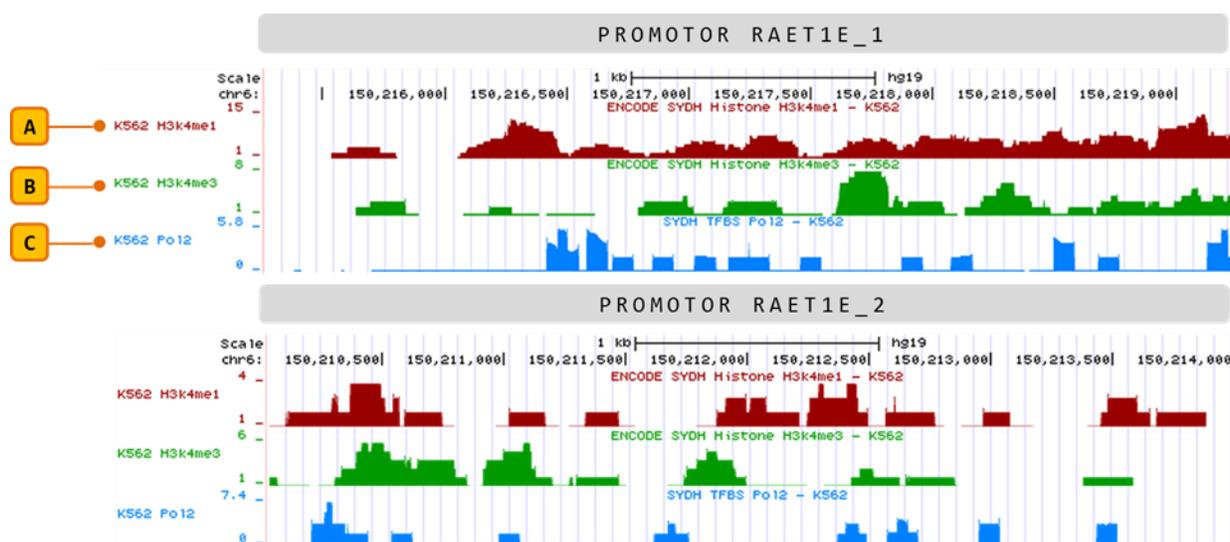


Figura 6. Predicción de la estructura de la cromatina para los promotores descritos del gen *RAET1E*. La figura muestra el estado de la cromatina en la región donde se han identificado los promotores 1 y 2 del gen *RAET1E*. Los elementos que se destacan corresponden a modificaciones químicas de la cromatina como metilaciones, las cuales influyen la expresión génica, así como también los sitios de unión para la ARN polimerasa II. En particular se destacan los siguientes elementos: **(A)** H3K4me1: mono-metilación de la lisina 4 de la histona 3. **(B)** H3K4me3: tri-metilación de la lisina 4 de la histona 3. **(C)** Po12: ARN Polimerasa II. Información colectada de UCSC Genome Browser⁵², disponible en: <http://genome.ucsc.edu/ENCODE>.

Resumiendo los antecedentes expuestos anteriormente podemos decir que:

- El gen *RAET1E* se caracteriza por ser el más divergente (43% de identidad de secuencia) dentro la familia de ligandos de NKG2D en humanos, lo que podría deberse a un rol distintivo de esta molécula.
- La expresión de RAET1E en algunos tipos de cáncer se ha identificado como un predictor de mal pronóstico.
- RAET1E se expresa en tumores de pacientes y líneas celulares de cáncer gástrico, mostrando un patrón de expresión heterogéneo.
- La expresión de RAET1E en líneas celulares de CG, frente al estímulo de *Helicobacter pylori*, no fue revertida al bloquear la molécula MyD88, corriente abajo de los TLR a diferencia de lo que ocurrió con los otros ligandos de NKG2D. Lo que podría deberse a que exista otra vía que regule su expresión.
- El promotor RAET1E_1 es el más activo transcripcionalmente y ha sido poco estudiado.
- Dado su carácter polimórfico, algunas de estas variaciones podrían localizarse en el promotor del gen *RAET1E*, alterando sitios de unión a factores de transcripción y por tanto modificando su expresión.
 - En un contexto de tumorigénesis tanto la disminución como la sobreexpresión de la molécula RAET1E favorecerían la evasión inmune permitiendo la progresión tumoral. Mientras que la disminución de la expresión de RAET1E se asociaría con una disminución de la citotoxicidad mediada por las células citolíticas, el aumento de esta molécula en el tiempo podría disminuir la expresión del receptor NKG2D en las células citolíticas afectando la función de las células NK.

HIPÓTESIS

“Variantes en el promotor *RAET1E* 1 en tumores de pacientes con adenocarcinoma gástrico modifican los sitios de unión para factores de transcripción”.

OBJETIVOS

1. OBJETIVO GENERAL

Identificar variantes en el promotor 1 de *RAET1E* que generen cambios en esta región reguladora y que puedan explicar la evasión de la respuesta inmune en tumores de pacientes con cáncer gástrico.

2. OBJETIVOS ESPECÍFICOS

1. Determinar la presencia de variantes en el promotor 1 del gen *RAET1E* en tumores de pacientes con cáncer gástrico.

2. Evaluar, a través de estudios bioinformáticos, aquellas variantes que generen cambios en los sitios de unión de factores de transcripción.

METODOLOGÍA

1. PACIENTES Y MUESTRAS BIOLÓGICAS

Este estudio se reclutaron 60 muestras de tumor de pacientes (20 mujeres y 40 hombres: edad $66,97 \pm 10,28$ años) diagnosticados con adenocarcinoma gástrico y pertenecientes al Hospital del Salvador, Región Metropolitana, Santiago. Los pacientes se sometieron a gastrectomía parcial o total, desde donde se obtuvo una muestra de tejido tumoral fresco (20-30 mg) el cual se almacenó en un tubo *ependorf* de 1,5 mL en solución PBS 1X (Phosphate Buffered Saline; stock 10X: NaCl 137 mM; KCl 2.7 mM; $\text{Na}_2\text{HPO}_4 \cdot 2 \text{H}_2\text{O}$ 10 mM; KH_2PO_4 2.0 mM; agua ultrapura; pH 7.4) y se mantuvo refrigerada a -20°C hasta su posterior procesamiento.

Las muestras empleadas en el estudio corresponden a una colaboración con el Dr. Patricio González, investigador de la Facultad de Medicina de la Universidad de Chile y que se colectaron previamente en el periodo 2013-2014.

El estudio cuenta con la aprobación del Comité de Ética en Investigación en Seres Humanos de la Facultad de Medicina de la Universidad de Chile y los respectivos consentimientos informados de quienes participaron de forma voluntaria (*Anexo 1*).

2. OBTENCIÓN DE ADN GENÓMICO

A partir de los 20 a 30 mg de tejido tumoral se procedió a extraer el ADN genómico de cada muestra empleando el kit comercial *QIAamp DNA Mini* (QIAGEN, EEUU) de acuerdo a las instrucciones del fabricante, considerando algunas modificaciones. El tejido tumoral se sometió en primer lugar a una digestión mecánica, para lo cual se empleó un micromortero hasta obtener una mezcla homogénea. Luego se continuó con la etapa de lisis, para lo cual se agregaron 180 uL de *buffer* ATL (QIAGEN, EEUU) y entre 15 a 40 uL de proteinasa K (QIAGEN, EEUU), dependiendo del tamaño del tejido tumoral. A continuación se incubaron los tubos a 56°C durante 3 e incluso 18 horas, mezclando vigorosamente de vez en cuando hasta lograr la completa degradación del tejido. Una vez finalizada la

incubación las muestras se trataron con ARNasa 100 mg/mL (US Biological, EEUU) por 5 minutos a temperatura ambiente. Las centrifugaciones realizadas en la etapa en que el lisado es transferido a la columna *QIAamp MinElute*, se realizaron a 10.000 rpm, requiriéndose en algunos casos prolongar el tiempo de centrifugación entre 3 a 5 minutos. Finalmente en la etapa de elución, se realizaron dos eluciones de 200 y 100 uL con el *buffer* AE (QIAGEN, EEUU), respectivamente.

Para los análisis posteriores se guardaron 20 uL de cada elución para cada una de las muestras.

3. ANÁLISIS CUANTITATIVO Y CUALITATIVO DEL ADN GENÓMICO

El ADN extraído de las muestras se sometió a diferentes mediciones para evaluar parámetros como cantidad, tamaño y estado de degradación, requeridos para estudios de secuenciación masiva.

3.1 Electroforesis en gel de Agarosa

La separación de fragmentos de ADN se realizó mediante geles de agarosa al 1% (US Biological, EEUU). Para la electroforesis se empleó el tampón TAE 1X (Tris-acetato 40 mM, EDTA 1 mM pH 8,0). Los geles se prepararon con TAE 1X y el agente intercalante de ADN bromuro de etidio (10 mg/mL, ThermoFisher Scientific, EEUU) en una concentración final 0.2-0.5 µg/mL. Como marcador de peso molecular se utilizó 1 kb DNA Ladder (Thermo Scientific, EEUU). Por último, las bandas de ADN genómico se visualizaron empleando un transiluminador (Viber Loumart).

3.2 Cuantificación del ADN genómico

3.2.1 Cuantificación Convencional

Se determinó la concentración del ADN genómico para cada muestra a través de espectroscopia UV, midiendo la absorbancia a 260 nm. Se evaluó la pureza del ADN, midiendo además la absorbancia a 280 nm, para establecer la relación A260/A280. Se consideraron como valores óptimos aquellos entre 1.8 y 2.0. Las mediciones se realizaron en el equipo CYTATION 3 (Bio-Tek, EEUU) utilizando en

cada caso 2 uL de muestra. Como blanco de la medición se utilizó el buffer de elución AE (QIAGEN, EEUU).

3.2.2 Ensayo de Cuantificación por PicoGreen

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

Preparación Buffer TE

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

Curva estándar de ADN

Como ADN estándar se utilizó el ADN Lambda (500 µg/mL, Invitrogen, EEUU). En primer lugar éste se diluyó a una concentración final de 5000 ng/mL en *buffer* TE 1X y luego se preparó la curva estándar siguiendo el protocolo que se detalla en la *tabla* 3.

Tabla 3. Protocolo preparación curva estándar para ensayo PicoGreen.

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

Reactivo PicoGreen

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

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

3.3 Electroforesis Capilar Paralela

La integridad y tamaño de los fragmentos de ADN presentes en cada muestra se evaluaron a través de una electroforesis capilar paralela. Se utilizó el kit comercial *High Sensitivity Genomic DNA* (Advanced Analytical, EEUU), para lo cual es necesario que las muestras de ADN genómico se diluyan a una concentración no mayor a los 10 ng/uL. Siguiendo las instrucciones del fabricante las muestras diluidas se analizaron en el bioanalizador *Fragment AnalyserTM* (Advanced Analytical, EEUU). Tras la corrida, el tamaño de los fragmentos de ADN se visualizó y analizó empleando el software PROSize 2.0 (Advanced Analytical, EEUU).

4. SECUENCIACIÓN DEL PROMOTOR DE RAET1E

Se realizó secuenciación masiva paralela empleando el equipo MiSeq (Illumina Inc., EEUU) y el kit *TruSeq Custom Amplicon* (Illumina Inc., EEUU). A continuación se detalla los pasos ejecutados desde el diseño hasta la etapa de secuenciación empleados para este estudio.

4.1 Diseño del proyecto de Secuenciación del promotor de RAET1E

El diseño consistió en la síntesis de pares de oligonucleótidos contenidos en el *Custom amplicon Tube* (CAT) para la generación de cada amplicón en la etapa de preparación de librerías. Para el diseño del CAT se utilizó el *software* facilitado por Illumina “*Design Studio*” disponible *on-line* en <http://designstudio.illumina.com/>. Se consideraron los siguientes parámetros en el diseño: (1) restringirse a la región del promotor de *RAET1E* comprendida entre las coordenadas genómicas del cromosoma 6: 150,217,262 a 150,219,767 (2506pb); (2) tamaño de los amplicones₁: 250 pb; (3) genoma de referencia: hg19; (4) indexado muestras: doble. Con estas especificaciones se envió la orden para que el Proyecto se diseñara por *Illumina*. En este caso en particular los amplicones requeridos fueron 12.

4.2 Preparación Librerías Genéticas

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

Como se muestra en la *figura 7*, este proceso contempló 7 etapas. En primer lugar, un grupo de oligonucleótidos específicos para el promotor de *RAET1E* se hibridó (corriente arriba y abajo) en el ADN genómico de cada una de las muestras. Luego, empleando un filtro de selección por tamaño, se removieron por completo los oligos no unidos al ADN. El tercer paso, consistió en conectar los oligos que hibridaron corriente arriba y debajo de la secuencia (extensión-ligación) blanco de

¹ Amplicón: Copia sintética de una región del ADN blanco, obtenida tras la etapa de amplificación de la preparación de librerías genéticas.

interés. Este paso permitió obtener las regiones blanco flanqueadas por secuencias requeridas en la siguiente etapa. Luego, se amplificó el producto del paso tres, usando partidores que permitieron agregar una secuencia índice en cada extremo (i5 e i7) así como adaptadores universales requeridos para la etapa de generación de grupos o clústers (P5 y P7). El quinto paso, consistió en limpiar los productos del PCR mediante esferas magnéticas (*Beads*), dejando fuera otros componentes de la reacción. A continuación, se procedió a normalizar la cantidad de las librerías. Esto con el fin de asegurar una representación equitativa de las librerías en el pool de muestras. Por último, las librerías ya normalizadas se agruparon combinando volúmenes iguales de cada librería. Para su secuenciación el grupo de librerías agrupadas se diluyó previamente en un *buffer* de hibridación (HT1) y denaturado por calor a 96°C durante 2 minutos.

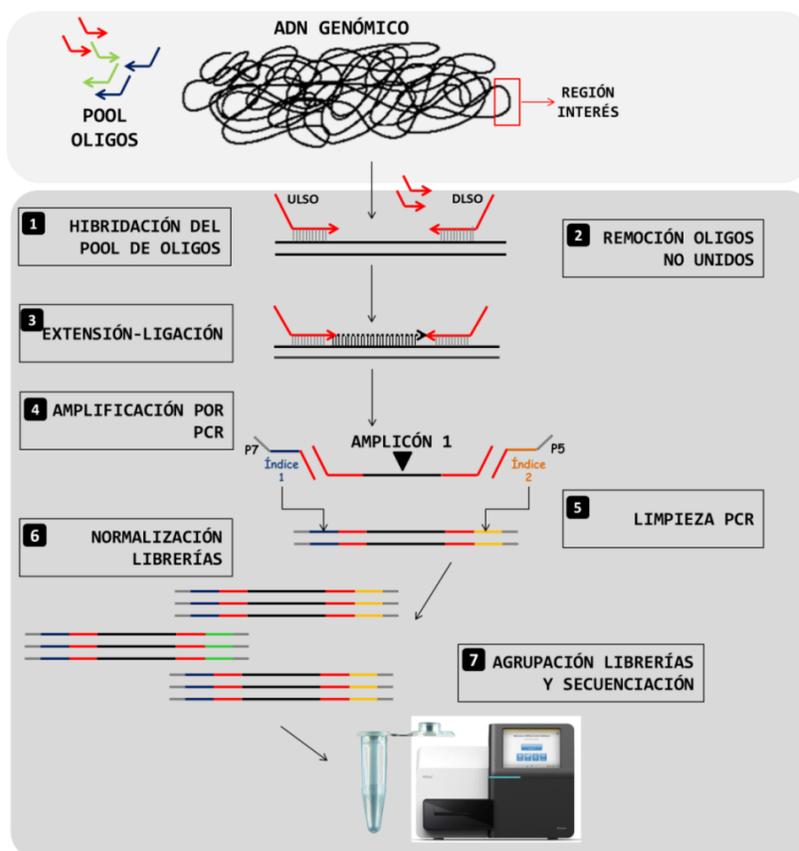


Figura 7. Representación gráfica de la preparación de librerías genéticas mediante el “Truseq Custom Amplicon”. La preparación de las librerías tiene 4 pasos fundamentales. En primer lugar, la hibridación de los oligos (ULSO y DSLO) a las a la región promotora de RAET1E (1). En segundo lugar, una etapa de extensión y ligación, donde las regiones de interés quedaron flanqueadas por secuencias de primers universales (3). En tercer lugar, una etapa de amplificación por PCR, le adiciona a los amplicones, índices (secuencias cortas de ADN) y adaptadores universales de secuenciación (4). Por último y en cuarto lugar, las librerías (compuestas por 12 amplicones para cada muestra de ADN) normalizadas se agruparon y secuenciaron en el equipo MiSeq (7). ULSO: Upstream Locus-Specific Oligo; DSLO: Downstream Locus-Specific Oligo.

En la preparación de la dilución del *grupo* de librerías (*DAL del inglés, Diluted Amplicon Library*) se utilizaron 7 uL del tubo PAL (*del inglés, Pooled Amplicon Library*) y 593 uL del buffer HT1.

Para confirmar que las librerías se amplificaron correctamente, se analizó una alícuota del control ACD1/ACP1 y de cada una de las 50 librerías en el bioanalizador *Fragment Analyser™* (Advanced Analytical, EEUU). Para ello se utilizó el kit

comercial *Standard Sensitivity NGS Fragment Analysis* (Advanced Analytical, EEUU), siguiendo las indicaciones proporcionadas por el fabricante.

4.3 Secuenciación masiva del promotor de RAET1E

Para la etapa de secuenciación en el equipo MiSeq, se utilizó el kit *MiSeq Reagent kit v.2* de 500 ciclos (Illumina Inc., EEUU), que permitió realizar lecturas de los amplicones en ambos sentidos. Este *kit* incluye el *cartridge* que contiene los reactivos necesarios para la secuenciación y corresponde al lugar físico donde se cargaron las librerías para ser secuenciadas. Luego este *cartridge* se cargó en el secuenciador junto con el archivo manifiesto de nuestro proyecto, el que especifica el genoma de referencia, así como las coordenadas genómicas de cada amplicón y el archivo hoja de muestras, que especifica la combinación de los índices para cada muestra. Ambos archivos son relevantes para el análisis posterior a la secuenciación.

Brevemente, las librerías de simple hebra se cargaron en el pocillo 17 del *cartridge* de secuenciación. Las librerías fluyeron a través de la celda de flujo donde hibridaron con los oligonucleótidos anti-sentido de P5 y P7 que son complementarios a los adaptadores de la librería. Como se mencionó antes, las librerías se diluyeron previo a ser cargadas en el secuenciador, esto para obtener grupos bien separados de idéntico producto (ejemplo: grupo amplicon 1, grupo amplicon 2, etc.) a partir de una única molécula de ADN. Los nucleótidos trifosfato marcados con fluoróforos se aplicaron en la celda de flujo y luego se excitaron por un láser. El espectro de emisión se guardó en el equipo *MiSeq*, y luego el nucleótido bloqueante, que inhibe futuras adiciones, se clivó permitiendo la adición de nuevos nucleótidos. De esta forma se permitió la adición de nuevos nucleótidos trifosfato y se secuenciaron cada uno de los fragmentos de ADN.

5. ANÁLISIS BIOINFORMÁTICO POST-SECUENCIACIÓN E IDENTIFICACIÓN DE VARIACIONES EN EL PROMOTOR DE RAET1E.

El flujo de trabajo requerido para identificar variaciones en el promotor de *RAET1E* en las 50 muestras estudiadas se resume en la *figura 9*. En primer lugar se obtuvieron los datos crudos de la secuenciación. Para ello un análisis primario realizado en tiempo real, convirtió los datos de imágenes a datos de intensidad para cada lectura², con la consiguiente transformación a datos de secuencia. Este análisis se realizó mediante el *software* RTA (*del inglés, Real Time Analysis*) integrado en el equipo MiSeq.

El diagrama muestra un archivo FASTQ con cuatro líneas de texto en un recuadro amarillo. A la derecha de cada línea hay un punto rojo conectado a una etiqueta en un cuadro amarillo: 'A' para la línea 1, 'B' para la línea 2, 'C' para la línea 3 y 'D' para la línea 4.

```
@SIM:FCX:1:15:6329:1045 1:N:0:2
TCGCACTCAACGCCCTGCATATGACAAGAGAATC
+
<>;##=><9=AAAAAAAAAAA9#:<#<:<<<???)#=#
```

Figura 8. Representación de los archivos FASTQ generados durante la secuenciación en el equipo MiSeq. En la figura se muestra el formato del archivo FASTQ, que se caracteriza por presentar 4 líneas: (A) El identificador, (B) Secuencia, (C) un signo más y (D) la puntuación de calidad en formato ASCII. Además el identificador (A) tiene su propio formato empezando con un @ Instrumento: ID Experimento: ID Celda de Flujo: Carril: Recuadro: X: Y: Número de Lectura: Indicador Filtro: 0: Número muestra.

Los datos de secuencia se almacenaron en un archivo FASTQ (*Figura 8*), que además de contener la letra para cada posición, indicó un valor de calidad para cada asignación de bases. Este valor numérico designado genéricamente como valor de calidad *Phred* (Q30 para la plataforma *Illumina*), nos indica la confianza con la que se predice una base en cada posición. En particular el Q30 indica que la probabilidad de asignar una base incorrectamente es de 1 en 1000⁵³.

²Lecturas: Secuencias cortas de ADN generadas a partir de la secuenciación por síntesis (SBS) en el secuenciador a partir de los amplicones.

Los análisis secundarios y terciarios de los datos se realizaron en la plataforma *on-line* de *Illumina* BaseSpace. Se subieron los 50 archivos FASTQ, y el archivo manifiesto a esta plataforma y se utilizó la aplicación “*TruSeq Amplicon app*” versión 1.1.0, que permitió precisamente evaluar regiones cortas de ADN en busca de variantes. Para ello el flujo de trabajo o *pipeline* de esta aplicación incluyó desde el alineamiento al llamado de variantes para el promotor de *RAET1E*.

Para la etapa de alineamiento se utilizó el algoritmo *Smith Waterman* en bandas, que realiza alineamientos locales para determinar regiones similares entre dos secuencias. En este caso entre las secuencias de nuestros amplicones y el genoma de referencia humano hg19. La particularidad de este algoritmo es que en vez de mirar en la secuencia completa, comparó segmentos de todos los tamaños posibles, con la restricción de que el tamaño máximo de INDELS (abreviación de inserciones y deleciones) fuese 25 pb. Luego, los datos de los alineamientos de las lecturas contra el genoma de referencia se almacenaron en archivos BAM (*del inglés, Binary Alignment/Map*) para cada muestra, los que pudieron ser visualizados usando “*Integrative Genomics Viewer*” (IGV) versión 2.3.

Posteriormente en la etapa de llamado o búsqueda de variantes, se utilizó el algoritmo *Somatic Variant Caller* (SVC) versión 3.2.2. Este algoritmo está especializado en el análisis de muestras de tumor que no se secuenciaron junto a las muestras de tejido no tumoral del mismo paciente, ajustándose a nuestros requerimientos. Así por cada variante SVC se anotó la secuencia de referencia y la variante, así como también la frecuencia y un valor de calidad (Q_{30}) indicando la confianza con la que la variante estuvo presente en la muestra.

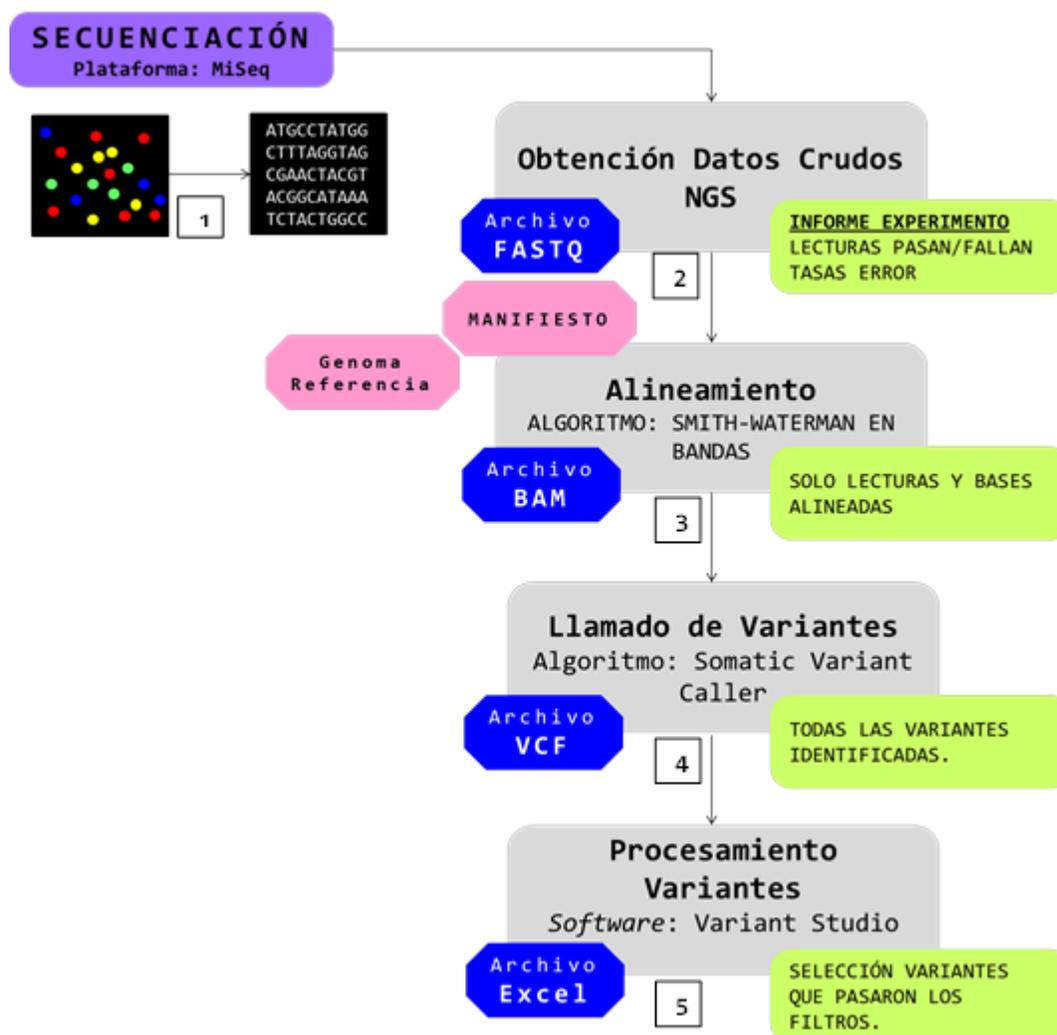


Figura 9. Flujo de trabajo empleado para el análisis de bioinformático de los datos obtenidos en la secuenciación del promotor de RAET1E. (1) (2) Tras la captura de imágenes y su traducción a secuencia, la información se almacenó en archivos FASTQ, los que además de contener la secuencia de los fragmentos pequeños de ADN o lecturas entrega información acerca de la calidad de los datos. **(3)** Luego este archivo se alineó, junto con el genoma de referencia y el archivo manifiesto, usando el algoritmo Smith Waterman en bandas. Del alineamiento, se generaron archivos BAM. **(4)** En la etapa de búsqueda o llamado de variantes se utilizó el algoritmo Somatic Variant Caller, generándose un archivo VCF, el que especificó las posiciones de las variantes. **(5)** Finalmente estos archivos se analizaron mediante el software Variant Studio para seleccionar solo aquellas variantes que pasaron todos los filtros y generar un archivo aplicable a los análisis posteriores.

Para la búsqueda de *Single Nucleotide Variant* (SNV) el algoritmo utilizado consideró cada posición en el genoma de referencia por separado, comenzando con las bases de la lectura alineada. Sólo se buscaron variantes cuya posición de la base tuviese una profundidad de cobertura igual o mayor a 10. Por último, en base al recuento entre bases mutadas y de referencia se calculó un valor de variante basado en un modelo Poisson. En este modelo se excluyen variantes si el valor Q es menor a 20, lo que significa tener una posibilidad de 1 en 100 de que ese SNV sea un falso positivo⁵⁴.

La información de las variantes identificadas en posiciones específicas del genoma se almacenó en archivos VCF (*del inglés, Variant Call Format*). Los archivos VCF de las 50 muestras se visualizaron y analizaron utilizando el programa Variant Studio 2.2. De esta forma se seleccionaron sólo aquellas variantes que pasaron todos los filtros (*ver tabla 4*), y se generó un archivo VCF final que se visualizó y convirtió a formato libro de Excel para los análisis posteriores. La *figura 10*, muestra un ejemplo de la tabla generada y los campos incluidos en ella.

Tabla 4. Filtro aplicados para la selección de variantes en el promotor de RAET1E.

NOMBRE FILTRO	DESCRIPCIÓN
Low Variant Freq	Variante que presenta una baja frecuencia, menor al 1%
Low GQ	Variante que presenta una calidad del genotipo menor a 30 (Escala 1 a 100)
R8	Aplicado para INDELS. El tamaño de las repeticiones adyacentes en el genoma de referencia es mayor a 8.
Low DP	Variante fue identificada en un sitio con baja cobertura. Menor a 10 lecturas.
Strand Bias	La variación identificada presenta diferencias de resultados entre la hebra <i>forward</i> en comparación a la hebra <i>reverse</i> .

Nº V	Muestra	Variant	Coordinate	Quality	GQX	Genotype	Alt Variant Freq	Read Depth	Alt Read Depth	Allelic Depths	dbSNP ID	Ancestral Allele	Allele Freq	Allele Freq Global Minor	Global Minor Allele	Allele Freq Amr	Allele Freq Asn	Allele Freq Af	Allele Freq Eur
1	CG58	C>T	150217232	10000	100	het	47.55	1592	757	833,757	rs564123677	C	0	0	0	0	0	0	0
2	CG17	C>T	150217284	10000	100	het	50.01	4259	2130	217,213	rs577576062	C	0	0	0	0	0	0	0
3	CG01	T>C	150217360	10000	100	hom	99.82	11829	11808	20,11808	rs9371544	C	52	47.76	T	50	85	33	41
	CG02	T>T	150217360	10000	100	het	39.90	8373	3341	5028,334	rs9371544	C	52	47.76	T	50	85	33	41
	CG06	T>C	150217360	10000	100	hom	99.70	12181	12144	34,12144	rs9371544	C	52	47.76	T	50	85	33	41
	CG07	T>C	150217360	10000	100	hom	99.89	11373	11361	12,11361	rs9371544	C	52	47.76	T	50	85	33	41
	CG08	T>T	150217360	10000	100	het	48.83	10123	4943	5180,494	rs9371544	C	52	47.76	T	50	85	33	41
	CG09	T>C	150217360	10000	100	hom	99.75	10431	10405	26,10405	rs9371544	C	52	47.76	T	50	85	33	41
	CG10	T>T	150217360	10000	100	het	50.23	7950	3933	3950,399	rs9371544	C	52	47.76	T	50	85	33	41
	CG13	T>T	150217360	10000	100	het	48.83	8747	4271	4471,427	rs9371544	C	52	47.76	T	50	85	33	41
	CG14	T>C	150217360	10000	100	hom	99.75	15101	15064	37,15064	rs9371544	C	52	47.76	T	50	85	33	41
	CG15	T>T	150217360	10000	100	het	75.82	5667	4297	1366,43	rs9371544	C	52	47.76	T	50	85	33	41
	CG16	T>C	150217360	10000	100	hom	99.78	5424	5412	12,5412	rs9371544	C	52	47.76	T	50	85	33	41
	CG17	T>C	150217360	10000	100	hom	99.78	5424	5412	12,5412	rs9371544	C	52	47.76	T	50	85	33	41
	CG18	T>T	150217360	10000	100	het	51.09	8240	4210	4024,421	rs9371544	C	52	47.76	T	50	85	33	41
	CG19	T>T	150217360	10000	100	het	52.50	8758	4598	4151,46	rs9371544	C	52	47.76	T	50	85	33	41
	CG21	T>T	150217360	10000	100	het	56.39	8643	4874	3767,487	rs9371544	C	52	47.76	T	50	85	33	41
	CG22	T>T	150217360	10000	100	het	48.20	12291	5924	6361,592	rs9371544	C	52	47.76	T	50	85	33	41
	CG23	T>T	150217360	10000	100	het	46.99	7400	3477	3911,348	rs9371544	C	52	47.76	T	50	85	33	41
	CG24	T>T	150217360	10000	100	het	46.39	12176	5649	6513,565	rs9371544	C	52	47.76	T	50	85	33	41
	CG28	T>T	150217360	10000	100	het	48.18	7304	3519	3777,352	rs9371544	C	52	47.76	T	50	85	33	41
	CG30	T>T	150217360	10000	100	het	50.30	3676	1849	1825,185	rs9371544	C	52	47.76	T	50	85	33	41
	CG32	T>T	150217360	10000	100	het	48.27	7706	3720	3986,372	rs9371544	C	52	47.76	T	50	85	33	41
	CG33	T>C	150217360	10000	100	hom	99.61	5644	5622	9,5622	rs9371544	C	52	47.76	T	50	85	33	41
	CG34	T>C	150217360	10000	100	hom	99.71	6545	6526	18,6526	rs9371544	C	52	47.76	T	50	85	33	41
	CG35	T>C	150217360	10000	100	hom	99.86	5704	5696	6,5696	rs9371544	C	52	47.76	T	50	85	33	41
	CG37	T>T	150217360	10000	100	het	36.43	3810	1388	2422,139	rs9371544	C	52	47.76	T	50	85	33	41
	CG38	T>C	150217360	10000	100	hom	99.76	6240	6225	15,6225	rs9371544	C	52	47.76	T	50	85	33	41
	CG40	T>T	150217360	10000	100	het	62.45	11623	7259	4362,726	rs9371544	C	52	47.76	T	50	85	33	41
	CG41	T>C	150217360	10000	100	hom	99.68	5570	5552	10,5552	rs9371544	C	52	47.76	T	50	85	33	41
	CG42	T>C	150217360	10000	100	hom	99.79	5258	5247	10,5247	rs9371544	C	52	47.76	T	50	85	33	41
	CG43	T>C	150217360	10000	100	hom	99.83	4729	4721	8,4721	rs9371544	C	52	47.76	T	50	85	33	41
	CG44	T>C	150217360	10000	100	hom	99.63	5100	5081	17,5081	rs9371544	C	52	47.76	T	50	85	33	41
	CG49	T>C	150217360	10000	100	hom	99.83	6915	6903	10,6903	rs9371544	C	52	47.76	T	50	85	33	41
	CG50	T>T	150217360	10000	100	het	52.09	7090	3633	3395,369	rs9371544	C	52	47.76	T	50	85	33	41
	CG51	T>C	150217360	10000	100	hom	99.78	6443	6429	14,6429	rs9371544	C	52	47.76	T	50	85	33	41
	CG52	T>C	150217360	10000	100	hom	99.81	7427	7413	13,7413	rs9371544	C	52	47.76	T	50	85	33	41

Figura 10. Archivo VCF final con las variantes que pasaron todos los filtros identificadas para el promotor de RAET1E. La figura representa un segmento de los datos totales que se obtuvieron para las variantes encontradas en cada uno de los pacientes.

6. ESTUDIO *IN SILICO* DE LA REGION DEL PROMOTOR DE RAET1E SECUENCIADA

6.1 Identificación De Sitios Putativos De Unión A Factores De Transcripción

Empleando la secuencia del genoma de referencia humano para los 2506 pb estudiados descargada de la base de datos NCBI⁴⁶, se realizó una predicción de los sitios de unión putativos para factores de transcripción. Para ello se utilizó el laboratorio virtual PROMO 3.0⁵⁵ que emplea la información depositada en la base de datos TRANSFAC⁵⁶ v8.3, disponible en: http://algggen.lsi.upc.es/cgi-bin/promo_v3/promo/promoinit.cgi?dirDB=TF_8.3. PROMO utiliza matrices de peso que representan los sitios de unión obtenidos a partir de datos experimentales, específicos para cada factor de transcripción⁵⁵. La tasa de disimilitud utilizada para la construcción de matrices fue 10%, es decir que los resultados se acotaron a aquellos cuyo porcentaje de disimilitud entre la secuencia ingresada (INPUT) y la secuencia consenso (Para el FT) fuese menor al 10%.

Además se realizó una segunda evaluación de los sitios de unión putativos para factores de transcripción. Se utilizó el software MatInspector⁵⁷, herramienta disponible en la plataforma virtual Genomatix (<http://www.genomatix.de/index.html>).

MatInspector también emplea matrices de peso de los sitios de unión para FTs, pero además asigna valores de calidad al localizar coincidencias dentro de la región promotora evaluada. Una coincidencia perfecta arroja un valor de matriz de similitud de 1.0⁵⁷. Los resultados obtenidos se filtraron seleccionando aquellos cuya puntuación de la matriz de similitud fuese mayor a 0.9. Además este software agrupa los factores similares y/o relacionados funcionalmente en familias de matrices, optimizando la entrega de resultados.

Por último, los resultados se acotaron a aquellas regiones donde se localizaron las variaciones de nucleótido simple (SNVs) para posteriormente evaluar el efecto de dichos cambios nucleotídicos en el promotor de *RAET1E*.

Todos los análisis se realizaron empleando el ensamblaje hg19/Build37 del genoma humano.

6.2 Evaluación Del Efecto De Los SNVs En Los Sitios Putativos De Unión A Factores De Transcripción

Con el objeto de evaluar si las variaciones identificadas produjeron un efecto en los sitios de unión de factores de transcripción, ya sea generando o suprimiendo sitios, se utilizó el *software* Variant Analysis incluido también en la plataforma virtual de Genomatix. Para ello fue necesario ingresar los ID de los SNVs (por ejemplo rs564123677). De manera intrínseca Variant Analysis utiliza el software MatInspector. Brevemente, para cada variante la secuencia genómica alrededor de la variante (aproximadamente 40 pb.) se analiza para localizar coincidencias con las matrices disponibles en la librería de matrices de Genomatix con las variantes alélicas insertadas en la secuencia de ADN. Luego los dos resultados son comparados y se entrega una lista de los factores de transcripción que se ven afectados por dicha variante.

En los casos en que no se tenía esta información, se evaluó el efecto de forma manual ingresando la secuencia con la variación y se analizó nuevamente la secuencia en búsqueda de sitios de unión a factores de transcripción empleando los *softwares* PROMO⁵⁵ 3.0 o MatInspector⁵⁷.

7. RECURSOS MATERIALES Y FINANCIAMIENTO

Esta tesis se desarrolló en las instalaciones de los laboratorios de Inmunovigilancia y Evasión Inmune, el Centro de InmunoBiotecnología (IBT) del Programa Disciplinario de Inmunología del ICBM, y el Laboratorio de Patología Molecular y Genómica del Cáncer del Centro de Investigación y Tratamiento del Cáncer & ICBM, Facultad de Medicina, Universidad de Chile, con los reactivos y el equipamiento allí disponible. La ejecución de la tesis fue financiada por los Proyectos Fondecyt N°1130330 (MCM), 1151446 (R.A) y FONDEF CA12I10023 (MCM) y D11I1029 (R.A) y Anillo en Ciencia y Tecnología ACT1115 (R.A).

RESULTADOS

OBJETIVO ESPECÍFICO 1: DETERMINAR LA PRESENCIA DE VARIANTES EN EL PROMOTOR DEL GEN RAET1E EN PACIENTES CON CÁNCER GÁSTRICO.

1. PACIENTES

El grupo de estudio contempló 50 de los 60 pacientes reclutados y diagnosticados con adenocarcinoma gástrico sometidos a gastrectomía total o parcial como parte de su tratamiento.

De estos 50 pacientes afectados por cáncer gástrico, 33 fueron hombres (66%) y 17 mujeres (34%), lo que da una relación hombre: mujer de 2:1.

El promedio de edad fue de 65,44 años y una desviación estándar de 10,28 años. El rango de edad fluctuó entre los 41 y 83 años, siendo de 46 – 78 años para las mujeres y de 41 – 83 años para los hombres.

En relación a la localización anatómica del tumor en la pared gástrica, se destaca que los dos lugares más frecuentes donde se presentaron las lesiones fueron la curvatura menor (38%) y la región antral (20%) (*Tabla 5*). En cuanto al tamaño del tumor, un 64% de los casos está por sobre los 5 cm de diámetro mayor, y dentro de este porcentaje un 22% presentó un diámetro mayor superior a los 10 cm. El rango del tamaño del tumor fue de 0,8 hasta los 20 cm.

Respecto a la etapa clínica de los tumores, considerando la clasificación TNM (*del inglés, Tumor, Node, Metastasis*), el 80% de los afectados presentó un adenocarcinoma de tipo avanzado, afectando todas las capas musculares del estómago hasta el tejido conectivo fuera de éste (T3) y en otros casos el tumor también había crecido en la membrana peritoneal o serosa (T4). Asimismo, la presencia de ganglios positivos se pesquisó en un 80% de los casos (*tabla 6*). De ellos, el 46% presentó invasión ganglionar en más de 15 linfonodos. En cuanto a metástasis a distancia, esta fue informada en el 10% de los pacientes, y en un 12% de los casos no se tuvo información al respecto (*tabla 7*).

Por último, en el 44% de los pacientes se detectó la presencia de la bacteria *H. Pylori* (*Tabla 5*), que como mencionamos anteriormente es un factor de riesgo importante para este tipo de cáncer.

Tabla 5. Características Clínico-patológicas de los pacientes con Cáncer Gástrico.

		Pacientes (N=50)	
Edad (años)		65.44 ± 10.28	
Sexo	Mujeres	17	
	Hombres	33	
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
= 0 > a 10 cm		11	
Estadío según el tamaño del tumor	T1	6	
	T2	4	
	T3	12	
	T4	28	
<i>H. pylori</i>	+	23	
	-	27	

NA: No aplica.

Tabla 6. Distribución de los casos según Invasión Gangliolar.

Invasión gangliolar	Nº Pacientes	%
Sin	9	18
Con	40	80
N.I	1	2
TOTAL	50	100

N.I: No informado.

Tabla 7. Distribución de los casos según Metástasis a distancia.

M. a Distancia	Nº Pacientes	%
M0	36	72
M1	5	10
MX	3	6
N.I	6	12
TOTAL	50	100

M0: No hay metástasis distante; M1: Presencia de metástasis distante; MX: No se pudo evaluar metástasis distante; N.I: No informado.

2. EVALUACIÓN CUANTITATIVA Y CUALITATIVA DEL ADN GENÓMICO

Con el objetivo de seleccionar sólo aquellas muestras que cumplieren con los requerimientos para la etapa de secuenciación, se analizó la calidad así como la cantidad de ADN genómico (ADNg) extraído a partir del tejido de 60 tumores de pacientes con adenocarcinoma gástrico.

En primer lugar, se evaluó el grado de degradación del ADNg a través de una electroforesis en gel de agarosa. Ello permitió clasificar las muestras en 5 grupos, desde el ADNg más integro al más degradado (Grupo I-V). Los tres primeros grupos dan cuenta de un ADN de mejor calidad, con menor signo de fragmentación y cuyo tamaño genómico se encuentra en el rango de los 6000 pb basados en el marcador de peso molecular (*Figura 11A*). Por lo tanto, tras esta etapa se descartaron 5 muestras pertenecientes a los Grupos IV y V.

Dado que las muestras del grupo II y III exhibían mayores patrones de degradación, se evaluó su calidad mediante una técnica más sensible como la electroforesis capilar paralela incluyendo como controles algunas muestras del grupo I y IV. La figura 11B muestra un electroferograma representativo para cada grupo, cuyo peak señala el tamaño promedio de los fragmentos de ADN que componen la muestra.

El análisis permitió corroborar que las muestras del grupo IV están mayormente degradadas, cuyo ADNg se compone de un conjunto de fragmentos cuyo rango de tamaños varía entre los 200 y 7000 pb. Sin embargo, el tamaño promedio de los fragmentos fue de 1770 pb. En el caso del grupo II el tamaño promedio de los fragmentos fue de 7514 pb y no hubo signos de degradación. Mientras que en el caso del grupo III algunas muestras presentaron un electroferograma similar al grupo IV, por lo cual fueron descartadas. Además, como criterio de selección se estableció que para la secuenciación se utilizarían sólo aquellas muestras cuyo tamaño promedio de los fragmentos fuese superior a los 3000 pb (peak mayor mostrado en el electroferograma).

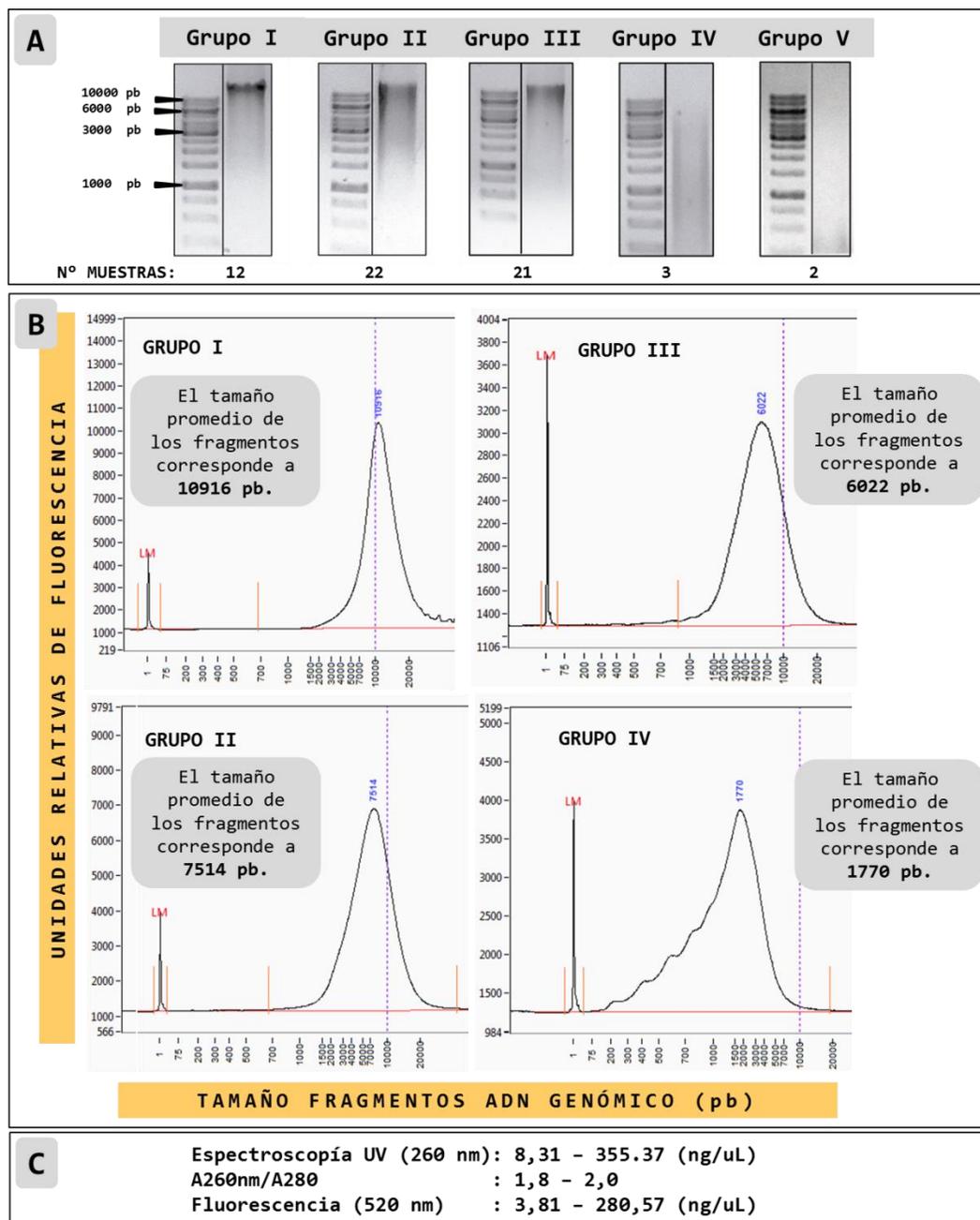


Figura 11. Evaluación Cuantitativa y Cualitativa del ADN genómico extraído de tumores de pacientes con adenocarcinoma gástrico. (A) Análisis de los fragmentos de ADNg mediante electroforesis en gel de agarosa. Se muestra una imagen representativa de cada grupo de acuerdo a su grado de degradación. Grupo I: ADN íntegro con una banda bien definida a la altura de los 10000 pb.; Grupo II: ADN sin banda definida, pero cuyo tamaño de fragmentos de concentra a la altura de los 10000 pb.; Grupo III: ADN con aparente grado de degradación y tamaño variable entre los 1500-10000 pb.; Grupo IV: ADN degradado y con fragmentos de tamaño menor a los 6000 pb.; Grupo V: Muestras cuyo ADN no fue visible en el gel de agarosa. (B) Análisis mediante electroforesis capilar paralela. Se muestra un electroferograma representativo para los grupos I, II, III y IV. En cada caso se observa una curva cuyo ancho en la base dependerá de cuan amplio sea el tamaño de los fragmentos que componen la muestra de ADNg. (C) Análisis cuantitativo del ADN genómico. Se muestra el rango de las concentraciones obtenidas por espectroscopía UV, mediante la medición de la absorbancia a 260 nm, y por emisión de fluorescencia a 520 nm para las 50 muestras utilizadas en este estudio.

EL otro parámetro relevante evaluado fue la cantidad de ADNg en cada una de las muestras. Primero se midió la concentración mediante espectroscopia UV a 260 nm y junto con ello se evaluó la pureza de las muestras mediante la relación de la absorbancia a 260 y 280 nm (A260/A280). Finalmente, se evaluó la concentración únicamente de ADN doble hebra mediante la emisión de fluorescencia a 520 nm.

Para la secuenciación, y en particular la preparación de las librerías genéticas, se requería tener una concentración inicial de 150-250 ng en un volumen de 10 uL, es decir, que por muestra se debía tener una concentración de 15-25 ng/uL (basados en la medición de concentración por fluorescencia). Entonces, el nuevo criterio de selección incluyó sólo muestras cuya concentración fue superior a 25 ng/uL, y cuya relación A260/A280 estuvo entre 1,8-2,0 (*figura 11C*). Por lo tanto en esta etapa se descartaron 2 muestras que pese a tener la calidad requerida no cumplían con el requisito de concentración. La *figura 12* resume en forma clara los pasos de selección que permitieron escoger las 50 mejores muestras para ser secuenciadas.

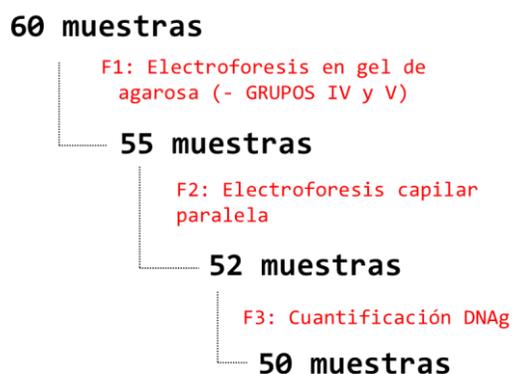


Figura 12. Flujo de trabajo empleado en la evaluación cualitativa y cuantitativa del ADNg de los pacientes en estudio. Las muestras se sometieron a varios filtros de selección de forma que cumpliesen con los requerimientos para ser utilizadas en la etapa de secuenciación. F1: Filtro de calidad del ADNg. Muestras clasificadas como degradadas según lo observado en gel de agarosa. F2: Filtro de calidad que descartó muestras cuyos fragmentos de ADN tuvieron un tamaño menor a 3000 pb. F3: Filtro de cantidad. Se descartó las muestras de ADNg cuya concentración fuese menor a 25 ng/μL. Abreviaciones: F1, F2 y F3, filtros 1, 2 y 3, respectivamente.

3. ANÁLISIS DE LA SECUENCIACIÓN DEL PROMOTOR DE RAET1E

La secuenciación del promotor *RAET1E-1* contemplaba el estudio de 2506 pb para cada uno de los pacientes. Durante el diseño del proyecto de secuenciación en la plataforma virtual *DesignStudio* se observó que la cobertura estimada para dicha región sería del 79%, dado que existía un segmento de 518 pb de baja especificidad para el diseño de los oligos. Por lo tanto, el diseño final abarcó un total de 2179 pb las que se cubrieron con 12 amplicones o subregiones (*Figura 13 C, D*). La generación de estos amplicones implicó la unión de los oligos diseñados a la hebra positiva o negativa del ADN. Como muestra la *figura 13D*, 7 amplicones se generaron a partir de la hebra positiva y 5 a partir de la hebra negativa lo que permitió obtener información de ambas hebras entre las coordenadas 150, 217,262 y 150,219,767 del cromosoma 6. La región excluida en el diseño se localizó entre las coordenadas 150,218,043 y 150,218,560 y por lo tanto no sabremos qué ocurre a nivel de variación en esta región del promotor.

Con el objetivo de evaluar la confiabilidad de los datos de secuenciación obtenidos, se analizó la profundidad de secuenciación, es decir las veces (X) que se leyó cada secuencia. Para ello se colectaron los datos de cobertura promedio (C.P) por cada amplicón. Este valor es fundamental dado que da confianza de los resultados obtenidos a partir de la zona secuenciada, permitiendo establecer la presencia de determinado nucleótido en una posición en particular.

Al analizar la C.P por cada amplicón se observó un patrón heterogéneo, que varió entre los 1102X y 8288X. Las regiones con menor cobertura fueron las correspondientes a los amplicones 9 (C.P:1102X) y 10 (C.P: 1144). Mientras que las regiones con mayores valores de cobertura correspondieron a los amplicones 2 (C.P: 8288) y 11 (C.P: 6100) (*Figura 13E*). Cuando se analizó las coberturas específicas por cada amplicón y paciente se concluyó entonces que cada posición de los fragmentos cortos de ADN secuenciados se leyó a lo menos 200 veces (Datos analizados pero no mostrados), lo que consideramos bueno basados en lo reportado en la literatura^{58,59}.

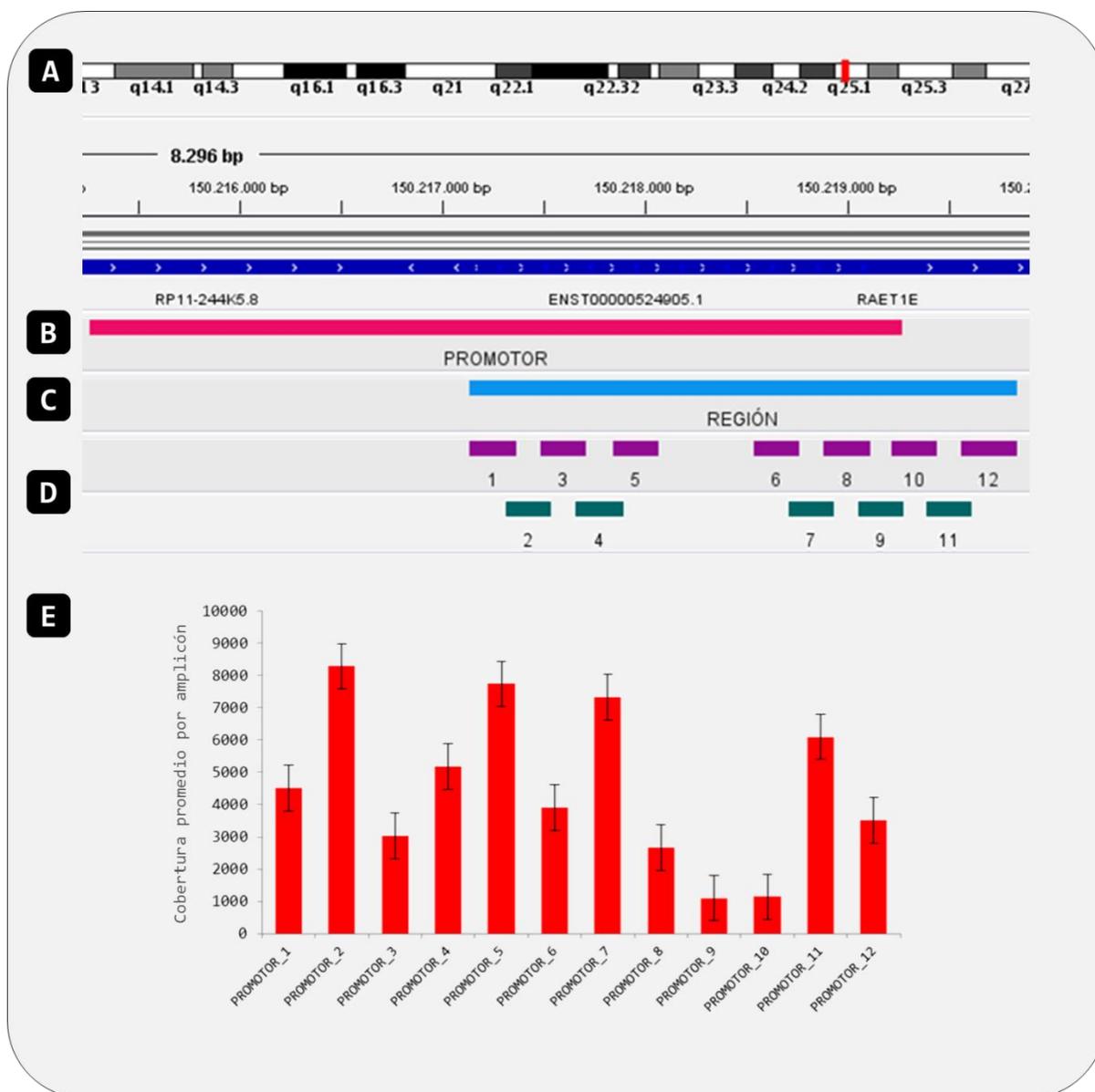


Figura 13. Esquema de los amplicones que mapean la región de estudio y análisis de su cobertura promedio. (A) Localización genómica de la región promotora de RAET1E. Se visualizan las bandas del cromosoma 6, y la línea roja señala el lugar específico que ocupa la región de estudio (q25.1) **(B)** Visualización de las 4001 pb que componen al promotor RAET1E-1. **(C)** Visualización de la región del promotor en estudio comprendida entre las coordenadas 150, 217,262 y 150,219,767 del cromosoma 6. **(D)** Localización de los 12 amplicones que mapean la región estudiada. 7 amplicones (**recuadros morados**) se generaron a partir de la hebra positiva y 5 amplicones (**recuadros verdes**) a partir de la hebra negativa. Además se observa la región del promotor que no fue cubierta correspondiente a un segmento de 518 pb localizado entre las coordenadas 150,218,043 y 150,218,560. **(E)** Gráfico que representa la cobertura promedio para cada amplicón.

4. IDENTIFICACIÓN DE VARIANTES EN EL PROMOTOR DE RAET1E

Una vez procesados los datos crudos de la secuenciación y obtenidos los archivos VCF (mencionado en apartado 5, Materiales y Métodos), se generó una tabla en formato Excel con todas las variantes identificadas para el promotor de *RAET1E*. Posteriormente, seleccionamos sólo aquellas variantes que pasaron todos los filtros bioinformáticos (ver tabla 4, apartado 5 Metodología). Se identificó un total de 49 variantes, el 86% de ellas del tipo SNV (del inglés, *Single Nucleotide Variation*) y el 14% restante fueron deleciones (Figura 14).

5 de los pacientes estudiados no presentó modificaciones de tipo SNVs, no obstante, se identificó en ellos una única deleción en la posición 150, 217,812.

Dado que los INDELS deben ser analizados mediante algoritmos bioinformáticos más avanzados, en el transcurso de esta investigación se decidió que el estudio se focalizaría en las variantes que se identificaron en un mayor porcentaje, es decir las variantes de nucleótido único (SNV).

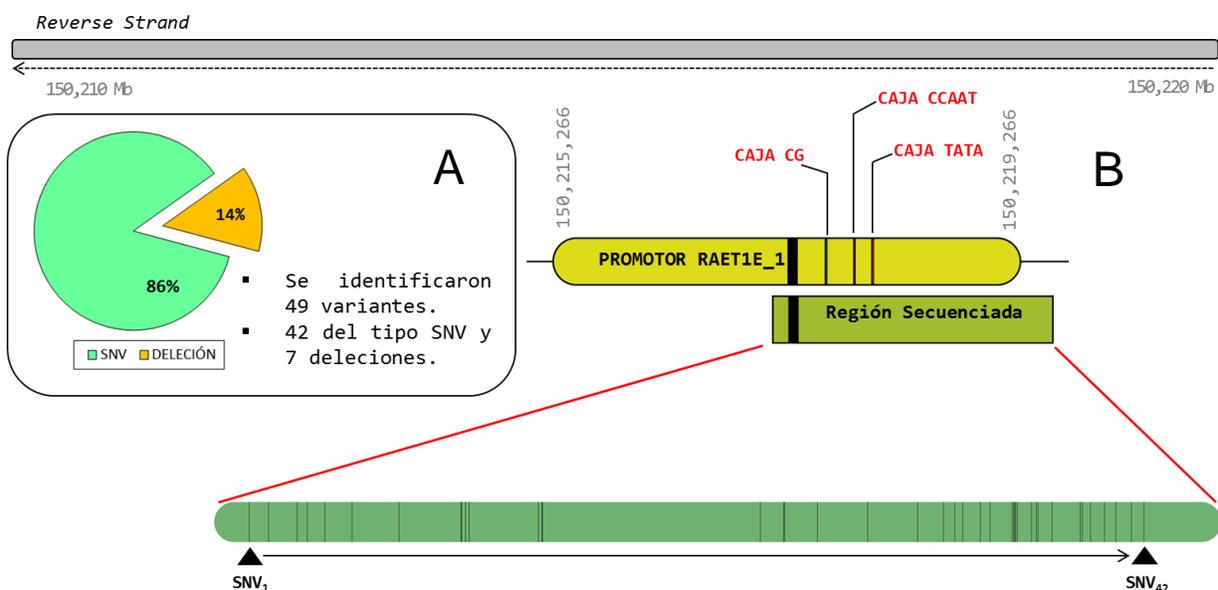


Figura 14. SNVs identificados en el promotor de *RAET1E* y su distribución en la región estudiada. (A) El gráfico muestra que se identificaron mayoritariamente variaciones de nucleótido simple (86%). **(B)** Estos 42 SNVs fueron ordenados de acuerdo a su coordenada genómica (versión CRCH37) y posteriormente se le asignó un número 1 al primero (SNV₁), sucesivamente hasta la variante 42 (SNV₄₂). Dentro de la región analizada encontramos elementos básicos de un promotor como las cajas CG, CCAAT y TATA.

Cada variante se ordenó de acuerdo a su coordenada en el genoma humano (hg19) de forma ascendente y se enumeró desde la variante N° 1 (SNV₁) hasta la variante N° 42 (SNV₄₂) (Columnas 1 y 2, *tabla 8*). De ellas, 19 no han sido descritas previamente en la base de datos dbSNP ni en COSMIC (Coordenadas destacadas en amarillo, *tabla 8*). Del total de SNVs identificados en el promotor de *RAET1E*, 13 variantes se detectaron en más de un paciente (recuadros azules) mientras que 19 fueron variantes únicas y distintivas del tumor (recuadros rojos).

La tabla se organizó de forma tal que los tumores de los pacientes (identificados como CGXX) se ordenaron de acuerdo al número de SNVs identificados en cada paciente y de acuerdo a la similitud en los patrones de SNVs por cada tumor. Es decir, desde los tumores con un menor a mayor número de variantes de nucleótido simple.

Así vemos que en los tumores de los pacientes CG04, CG25, CG29, CG36 y CG46 no se detectó ninguno de los 42 SNVs identificados en este estudio. Entre estas muestras no existen característica clínico-patológica en común, pero como se mencionó anteriormente, en ellos se identificó una delección en la misma coordenada genómica. Luego en el tumor CG53, se identificó un único SNV (SN₂₁) el cual fue exclusivo para este tumor, no encontrándose en el resto de los pacientes. El 84% de los tumores presentó entre 3 a 13 variaciones de nucleótido simple, siendo la moda la presencia de 3 y 4 SNVs. Interesantemente, 2 tumores (CG45 y CG02) presentaron un patrón distintivo de 8 SNVs en común, que no se identificaron para los otros tumores. Si bien ambos tumores comparten la misma localización anatómica (antral) no comparten otra característica clínica.

Asimismo identificamos un patrón de 3 SNVs ya descritos previamente (SNV₃, SNV₁₁ y SNV₄₂) presentes en 36 de los 50 tumores. Llamativamente identificamos dos SNVs no descritos previamente: SNV₄ y SNV₂₀, presentes en 5 y 10 tumores respectivamente. La presencia simultánea de este patrón solo se identificó en un tumor (CG58).

En el contexto de la genómica y el cáncer, podemos identificar en los individuos de una población variaciones germinales que podrían asociarse con una mayor o menor susceptibilidad a padecer esta enfermedad, o bien variaciones

somáticas desarrolladas durante el proceso tumorigénico y que resultan relevantes tanto para el entendimiento de los mecanismos implicados en cáncer así como desde el punto de vista terapéutico⁶⁰. Para las 42 variantes identificadas determinamos la frecuencia alélica a partir de la fracción de lecturas que incluyeron la variante para una determinada muestra. Basados en este parámetro así como la información disponible en la base de datos dbSNP observamos los dos tipos de variaciones mencionadas anteriormente germinales y somáticas (*Tablas 9 y 10*).

En este estudio identificamos 25 SNVs de tipo germinal, cuyas frecuencias alélicas estimadas fueron de 50% para heterocigotos y 100% para homocigotos. De ellos 2 no han sido reportados previamente: los SNV5 y SNV9, identificados en los tumores CG49 y CG55 respectivamente (*Tabla 9*). Los 22 SNVs restantes, fueron identificados como variantes de tipo somáticas, y su frecuencia varió entre 1,2 a 18% (*Tabla 10*). El límite de detección de variantes usando la tecnología MiSeq y el algoritmo Somatic Variant Caller fue determinado previamente, identificando que se pueden identificar variantes desde una frecuencia igual o mayor al 1% (*Datos correspondientes al estudio realizado en el laboratorio de Patología Molecular y Genómica del Cáncer, Facultad de Medicina, Universidad de Chile*).

Tabla 9. Frecuencias Alélicas de las variantes germinales identificadas en el promotor de RAET1E.

VARIANTES DE TIPO GERMINALES											
Nº SNV	Coordenada Genómica (CRCH37)	Nº Tumores de Pacientes	Tumor	GENOTIPO	Variante ^a	Frecuencia Alélica ^b	ID dbSNP	Frecuencias Poblacionales ^c			
								F.A Amr	F.A Asn	F.A Af	F.A Eur
3	150,217,360	17 23	*	Heterocigotos Homocigotos	T>T/C T>C/C	50,76 99,78	rs9371544	48,99	87,00	32,75	41,95
11	150,217,811	17 23	*	Heterocigotos Homocigotos	C>C/T C>T/T	52,32 99,78	rs6935829	48,99	87,00	32,75	41,95
42	150,219,627	22 14	*	Heterocigotos Homocigotos	G>G/A G>A/A	49,97 99,37	rs12528243	37,46	33,04	16,81	26,24
19	150,218,753	2		CG02 CG45	Heterocigotos	C>C/A	rs3964083	4,18	0	32,83	0,10
20	150,218,887	2		CG02 CG45	Heterocigotos	G>G/A	rs7761039	4,18	0	32,83	0,10
23	150,219,121	2		CG02 CG45	Heterocigotos	G>G/T	rs2342771	4,18	0	31,92	0,10
29	150,219,281	2		CG02 CG45	Heterocigotos	T>T/C	rs7750945	2,18	0	21,18	0,20
32	150,219,326	2		CG02 CG45	Heterocigotos	G>G/C	rs7761734	2,88	0	21,03	1,00
		9	*		G>G/A	52,91		11,24	53,67	6,28	15,11
35	150,219,381	2		CG02 CG45	Heterocigotos	G>G/A	rs7761879	4,18	0	34,04	1,00
39	150,219,521	2		CG02 CG45	Heterocigotos	A>A/G	rs28521726	2,88	0	21,18	1,00
40	150,219,551	2		CG02 CG45	Heterocigotos	C>C/T	rs28580145	2,88	0	21,18	1,00
6	150,217,507	1		CG02	Heterocigotos	G>G/A	rs568178294	0	0	1,70	0
8	150,217,657	1		CG45	Heterocigotos	C>C/T	rs80328924	0,29	0	4,24	0
13	150,218,006	1		CG07	Heterocigotos	G>G/T	rs7755896	1,40	0	5,80	0
14	150,218,016	1		CG07	Heterocigotos	G>G/T	rs7755900	1,40	0	5,80	0
15	150,218,017	1		CG07	Heterocigotos	C>C/T	rs7770964	1,40	0	5,80	0
17	150,218,662	1		CG23	Heterocigotos	T>T/C	rs191956816	2,90	0	0,08	1,09
24	150,219,142	1		CG52	Heterocigotos	T>T/A	rs150896420	1,00	0	2,90	0
25	150,219,189	1		CG02	Heterocigotos	C>C/T	rs2342772	1,59	0	12,63	0
36	150,219,456	1		CG45	Heterocigotos	C>C/T	rs28564789	2,74	0	19,52	1,00
1	150,217,232	1		CG58	Heterocigotos	C>C/T	rs564123677	0	0	0	0
2	150,217,284	1		CG17	Heterocigotos	C>C/T	rs577576062	0	0	0	0
21	150,219,021	1		CG53	Heterocigotos	C>C/T	rs113083455	0,58	1,19	6,2	1,19
5	150,217,435	1		CG49	Heterocigotos	A>A/G		0	0	0	0
9	150,217,799	1		CG55	Heterocigotos	T>T/A		0	0	0	0

^a Cada variante muestra el alelo mayor y a continuación el genotipo identificado en cada caso. Ej.: T>T/C; donde T es el alelo mayor y T/C el genotipo.

^b La frecuencia alélica fue determinada a partir de los reads o lecturas obtenidas para cada alelo. Se informa la frecuencia del alelo alternativo.

^c Las frecuencia poblacionales fueron obtenidas de la base de datos del proyecto 1000 Genomes (disponible en www.1000genomes.org/). Abreviaciones: **F.A**: Frecuencia Alélica; **Amr**: Amerindios; **Asn**: Asiáticos; **Af**: Africanos. Dado que en los SNVs 3, 11, 42 y 32 el número de tumores que presentan la variante es mayor a 5, no se han puesto en la tabla 9 y la frecuencia alélica corresponde al promedio de todas las muestras para cada variante (*más detalles Anexo X*). Los SNVs no descritos se han resalta en gris.

Tabla 10. Frecuencias Alélicas de las variantes somáticas identificadas en el promotor de RAET1E.

VARIANTES DE TIPO SOMÁTICAS					
Nº SNV	Coordenada Genómica (CRCH37)	Nº Tumores de Pacientes	Tumor	Variante ^a	Frecuencia Alélica ^b
4	150,217,387	5	CG59	T>T/C	1,27
			CG40		1,36
			CG22		1,41
			CG19		1,53
			CG58		2,32
10	150,217,801	10	CG33	T>T/C	1,64
			CG60		1,53
			CG43		1,66
			CG13		1,5
			CG39		1,79
			CG27		1,82
			CG50		1,72
			CG09		1,76
			CG30		2,5
CG58	2,46				
7	150,217,633	1	CG21	T>T/A	1,86
12	150,217,821	1	CG06	T>T/C	18,44
16	150,218,600	1	CG21	A>A/T	2,66
18	150,218,665	1	CG57	C>C/T	4,92
22	150,219,090	1	CG22	A>A/G	2,09
26	150,219,215	1	CG40	T>T/C	2,44
27	150,219,274	1	CG55	C>C/A	3,05
28	150,219,278	1	CG55	G>G/A	2,68
30	150,219,283	1	CG57	T>T/C	3,98
31	150,219,288	1	CG32	A>A/G	2,12
33	150,219,339	1	CG59	T>T/C	2,11
34	150,219,343	1	CG34	A>A/G	2,04
37	150,219,462	1	CG21	T>T/A	1,55
38	150,219,484	1	CG21	T>T/C	1,8
41	150,219,593	1	CG57	T>T/C	2,76

^a Cada variante muestra el alelo mayor y a continuación el genotipo identificado en cada caso, igual como se explicó en tabla 9.

^b La frecuencia alélica fue determinada a partir de los reads o lecturas obtenidas para cada alelo. Se informa la frecuencia del alelo alternativo.

OBJETIVO ESPECÍFICO 2: EVALUAR, A TRAVÉS DE ESTUDIOS BIOINFORMÁTICOS, AQUELLAS VARIANTES QUE GENEREN CAMBIOS RELEVANTES EN EL PROMOTOR DE RAET1E.

La existencia de variantes regulatorias que afecten la transcripción ha sido bastante reportada y en la actualidad se sabe que tendrían una importante función en las enfermedades humanas⁶¹. Por ello, evaluamos la presencia tanto de elementos reguladores basales como distales en la región del promotor objeto de estudio. En particular, acotando los análisis a las regiones donde se identificaron los SNVs.

1. ELEMENTOS REGULADORES DE TIPO BASAL PRESENTES EN EL PROMOTOR DE RAET1E.

Dentro de los elementos basales de un promotor identificados en la región secuenciada (2696 pb), distinguimos el sitio de inicio de la transcripción (TSS, *del inglés Transcription Start Site*), que de acuerdo a la información disponible en la base de datos EPD se localizaría dentro de una región comprendida aproximadamente por 60 pb (chr6:150,217,256-150,217,315) (*figura 14B*). EPD, es una base de datos que colecta información de promotores validados experimentalmente, y que además provee información de algunos elementos reguladores presentes en el promotor estudiado⁵⁰.

Los elementos basales de regulación que se encontraron en el promotor 1 de *RAET1E* serían, *la caja CG*, la caja CCAAT y la caja TATA. Estos elementos se localizarían entre las coordenadas 150, 217,602 y 150, 218, 000 (*figura 14B*).

2. ANÁLISIS IN SILICO DE SITIOS PUTATIVOS DE UNIÓN PARA FACTORES DE TRANSCRIPCIÓN (FT).

Para predecir los sitios de unión a FTs, con PROMO 3.0 y MatInspector, se utilizó como secuencia de entrada, la secuencia disponible en la base de datos NCBI comprendida entre las coordenadas 150, 217,139 y 150, 219,834 del cromosoma 6 genoma de referencia hg19. En el caso de PROMO 3.0 se analizó tanto la hebra positiva como la hebra negativa. Mientras que en MatInspector sólo se usó la hebra positiva, dado que de forma automática analiza la hebra negativa. En ambos casos,

los resultados se acotaron a buscar sitios putativos de unión a factores de transcripción en las regiones variables identificadas en nuestro estudio, es decir, las posiciones genómicas de los 42 SNVs, que corresponden a los sitios que potencialmente podrían modificar el reconocimiento del respectivo FT.

2.1 Predicción de los Sitios de unión a FTs con PROMO 3.0

Cuando se analizó la hebra negativa en dirección 3' a 5' (*Tabla 12*), hubo una mayor cantidad de sitios de unión putativos que al analizar la hebra positiva en dirección 5' a 3' (*Tabla 11*). Lo anterior podría deberse al hecho de que el gen RAET1E se transcribe a partir de la hebra negativa.

En la localización genómica de 29 SNVs se identificaron sitios de unión para factores de transcripción con una similitud de secuencia mayor o igual al 90%. Del análisis de la hebra positiva, se encontró un sitio de unión para el factor de transcripción general TFIID en la posición correspondiente al SNV₁₆ cuya similitud con la secuencia consenso fue del 100% y con valores esperados (E-values) bajos. Este valor (E-value) indica la expectativa de encontrar esta coincidencia por azar y considera la equiprobabilidad de los 4 nucleótidos (RE equally) así como la misma frecuencia de los 4 nucleótidos (RE query), por lo tanto valores bajos (menores a 1) indican mayor confiabilidad en la predicción.

Al analizar la hebra negativa (*tabla 12*) se encontraron sitios de unión con una coincidencia del 100% para los factores de transcripción TFII-I (SNV₁₃), IRF-2 (SNV₂₂), NF-1 (SNV₃₂) y GR- β (SNV₁₆).

Dentro de los sitios de unión putativos para factores de transcripción encontrados, destacamos sitios de unión cuyo porcentaje de similitud con la secuencia consenso está en 96-98%, y cuya función en el contexto del sistema inmune y el cáncer podrían ser relevantes. Entre ellos encontramos los factores STAT4 (SNV₃), IRF-1 (SNV₁₃) y HIF-1 (SNV₂₃).

Tabla 11. Sitios putativos de unión a FTs, identificados en la hebra positiva de la región promotora de RAET1E (Predicción realizada mediante PROMO 3.0).

FACTOR DE TRANSCRIPCIÓN	POSICIÓN EN SEC. BLANCO ^a	LOCALIZACIÓN SITIOS UNIÓN ^b	SECUENCIA	% DISIMILITUD ^c	RE EQUALLY ^d	RE QUERY ^e	SNV ^f	COORDENADA GENÓMICA ^g
GR- α	222	219-223	CCTTA	6,26	5,27	5,25	3	150217360
TFIID	495	495-501	TCTCAA	8,01	2,97	2,67	7	150217633
GR- α	868	865-869	AAAGG	0,21	10,53	10,53	13	150218006
c-Est-1	868	868-874	GGGGAA	6,94	0,99	1,00	13	150218006
TFIID	1462	1458-1464	TTTTAA	4,01	1,48	1,26	16	150218600
TFIID	1462	1459-1465	TTTAAA	0,00	1,48	1,27	16	150218600
EBF	1524/1527	1524-1534	TGCCCTGGGCT	5,03	0,03	0,03	17/18	150218662/665
GR- α	1527	1527-1531	CCTGG	8,07	10,53	11,10	18	150218665
p53	1615	1609-1615	AAGGCC	7,64	0,99	1,08	19	150218753
PAX-5	1615	1609-1615	AAGGCC	4,01	1,48	1,60	19	150218753
GR- α	1749	1746-1750	AGAGG	0,21	10,53	10,53	20	150218887
NFI/CTF	1883	1883-1890	CCAACCC	7,01	0,99	1,00	21	150219021
RAR- β	1883	1881-1890	CCCCAACCC	9,64	0,29	0,30	21	150219021
TFII-I	1952	1951-1956	CACTCC	6,58	1,31	1,35	22	150219090
PR A	1983	1983-1989	GGGTGT	9,74	1,48	1,47	23	150219121
XBP-1	2004	1999-2004	AGCCAT	9,79	2,63	2,63	24	150219142
TFII-I	2004	2001-2006	CCATCC	9,51	9,87	10,06	24	150219142
c-Jun	2077	2077-2083	TGACTGC	7,10	0,99	0,99	26	150219215
GR- α	2150	2148-2152	AGAGG	0,21	10,53	10,53	31	150219288
TFII-I	2150	2146-2151	GGAGAG	1,82	0,66	0,70	31	150219288
AP2- α	2150	2148-2153	AGAGGC	3,23	0,66	0,70	31	150219288
GR- α	2201	2199-2203	CCTGT	0,00	10,53	10,53	33	150219339
c-Myb	2205	2204-2211	CAACTGTC	3,47	0,16	1,16	34	150219343
TFIID	2346	2346-2352	TTTAGCA	7,08	0,16	0,15	38	150219484
XBP-1	2455	2450-2455	TGGCAT	6,48	1,32	1,35	41	150219593
ENKTF-1	2455	2450-2457	TGGCATAG	5,69	0,99	1,03	41	150219593

a Posición del SNV en la secuencia total analizada a partir de la secuencia en genoma de referencia (Total 2696 pb).

b Localización sitio unión putativo para determinado factor de transcripción.

c El porcentaje de Disimilitud corresponde al porcentaje de diferencia de la secuencia consenso para determinado factor de transcripción versus la secuencia en la secuencia obtenida desde el genoma de referencia. secuencia analizada. Valores iguales a 0 indican 0% de disimilitud, es decir, una similitud del 100%.

d y e RE EQUALLY Y RE QUERY, corresponden a mediciones de la confiabilidad de las predicciones. Los valores esperados calculados consideran en el primer caso, un modelo matemático que calcula la equiprobabilidad de los 4 nucleótidos, y en el segundo caso un modelo con la misma frecuencia de los nucleótidos para la "secuencia problema" analizada.

f Se indica el SNV que alteraría el sitio de unión putativo para determinado factor de transcripción. Ej: 41, corresponde al SNV número 41.

g Coordenada genómica del SNV identificado.

Tabla 12. Sitios putativos de unión a FTs, identificados en la hebra negativa de la región promotora estudiada de RAET1E, (Predicción realizada mediante PROMO 3.0).

FACTOR DE TRANSCRIPCIÓN	POSICIÓN EN SEC. BLANCO ^a	LOCALIZACIÓN SITIOS UNIÓN ^b	SECUENCIA	% DISIMILITUD ^c	RE EQUALLY ^d	RE QUERY ^e	SNV ^f	COORDENADA GENÓMICA ^g
Pax-5	146	143-149	ACGGCCC	4,01	1,49	1,6	2	150217284
p53	146	143-149	ACGGCCC	7,64	0,99	1,01	2	150217284
GR-β	222	221-225	AATAC	5,04	5,27	4,97	3	150217360
STAT4	222	219-224	GGAATA	2,94	3,95	3,95	3	150217360
c-Est-1	222	217-223	CGGGAAT	9,59	0,5	0,5	3	150217360
GR-β	297	294-298	TCATT	0,84	10,53	9,68	5	150217435
NF-1	297	297-304	TTGGTACC	7,69	0,33	0,32	5	150217435
RXR-α	369	363-369	GTCACCC	5,94	0,99	1,06	6	150217507
c-Myb	369	367-374	CCCAGTTC	5,14	0,41	0,4	6	150217507
TFII-I	519	518-523	GGACAA	9,51	9,87	10,06	8	150217657
TFII-I	868	864-869	CTTTCC	0,00	1,98	2,01	13	150218006
STAT4	868	864-869	CTTTCC	2,94	3,95	3,95	13	150218006
c-Est-1	868	865-871	TTTCCC	6,94	0,99	1	13	150218006
IRF-1	868	865-873	TTTCCCCTT	3,35	0,09	0,09	13	150218006
NF-AT1	868	861-869	TGCTTTCC	6,25	0,05	0,05	13	150218006
NF-AT1	868	868-876	CCCCTTCC	9,18	0,3	0,3	13	150218006
RelA	868	862-872	GTCTTTCCCCT	8,60	0,03	0,03	13	150218600
NF-kappaB1	868	861-871	TGCTTTCCCCT	8,08	0,04	0,05	13	150218006
c-Est-1	878	872-878	TTTCCC	6,94	0,99	1	14	150218016
c-Est-1	878	878-884	CGGGAAG	5,43	0,5	0,53	14	150218016
IRF-1	878	872-880	TTTCCCCTG	7,04	0,17	0,17	14	150218016
STAT1β	878/879	871-880	CTTCCCCTG	8,70	0,3	0,3	14/15	150218016/017
GR-β	1462	1459-1463	AAATT	0,00	5,27	4,70	16	150218600
p53	1615	1613-1619	GGGCCGT	7,64	0,99	1,08	19	150218753
TFII-I	1749	1745-1750	CTCTCC	1,82	0,66	0,7	20	150218887
NFI/CTF	1883	1881-1888	GGGGTTGG	7,01	0,99	1	21	150219021
Pax-5	2188	2184-2190	TTGGCCC	9,55	1,98	2,2	32	150219021
p53	2188	2184-2190	TTGGCCC	6,10	0,82	0,93	32	150219326
RXR-α	1883	1882-1888	GGGGTTGG	2,73	1,15	2,01	21	150219021
RAR-β	1883	1881-1890	GGGGTTGGGG	9,64	0,29	0,3	21	150219021
GR-α	1952	1952-1956	TGAGG	6,26	5,27	5,25	22	150219090
STAT4	1952	1947-1952	GGAAGT	2,94	3,95	3,95	22	150219090
IRF-2	1952	1949-1954	AAGTGA	0,00	0,66	0,62	22	150219090
HIF-1	1983	1978-1986	ACGTGCCCA	2,02	0,17	0,18	23	150219121
GR-α	2004	2002-2006	GTAGG	8,07	10,53	11,11	24	150219142
RXR-α	2136/2140	2136-2142	GAGACCC	4,87	0,66	0,7	27/28	150219274/278
GR-α	2150	2146-2150	CCTCT	0,21	10,53	10,53	31	150219288
TFII-I	2150	2147-2152	CTCTCC	1,82	0,66	0,7	31	150219288
RXR-α	2188	2188-2194	CCCACCC	5,27	0,82	0,88	32	150219326
NF-1	2188	2184-2191	TTGGCCCA	0,00	0,16	0,18	32	150219326
FOXP3	2205	2204-2209	GTTGAC	4,76	3,95	3,77	34	150219343
c-Myb	2205	2200-2207	GACAGTTG	3,47	0,16	0,16	34	150219343
GR-α	2243	2242-2246	GCAGG	8,07	10,53	11,11	35	150219381
TFIID	2324	2320-2326	TTTAAGA	4,01	1,48	1,26	37	150219462
GATA-1	2455	2454-2459	TATCGA	2,18	5,27	5,11	41	150219593
GATA-2	2455	2450-2458	ACCGTATCG	6,67	0,33	0,32	41	150219593

a Posición del SNV en la secuencia total analizada a partir de la secuencia en genoma de referencia (Total 2696 pb).

b Localización sitio unión putativo para determinado factor de transcripción.

c El porcentaje de Disimilitud corresponde al porcentaje de diferencia de la secuencia consenso para determinado factor de transcripción versus la secuencia en la secuencia obtenida desde el genoma de referencia. secuencia analizada. Valores iguales a 0 indican 0% de disimilitud, es decir, una similitud del 100%.

d y e RE EQUALLY Y RE QUERY, corresponden a mediciones de la confiabilidad de las predicciones. Los valores esperados calculados consideran en el primer caso, un modelo matemático que calcula la equiprobabilidad de los 4 nucleótidos, y en el segundo caso un modelo con la misma frecuencia de los nucleótidos para la "secuencia problema" analizada.

f Se indica el SNV que alteraría el sitio de unión putativo para determinado factor de transcripción. Ej: 41, corresponde al SNV número 41.

g Coordenada genómica del SNV identificado.

2.2 Predicción de los Sitios de unión a FTs con MatInspector

La predicción de los sitios de unión a FT_s con MatInspector fue más acotada y restrictiva. En este caso el valor que nosotros evaluamos fue la matriz de similaridad, cuyo valor sobre los 0,8 se considera una buena predicción. Un valor de 1,00 indica una coincidencia total con la secuencia consenso de la matriz analizada (*tablas 14 y 15*). En nuestro análisis encontramos que se obtuvo un valor de 1,00 para los sitios de unión putativos de MZF1 (SNV₁₃), ZKSCAN3 (SNV₂₄), ARNTL (SNV₃₄) y FOXP1 (SNV₄₀) (*Tabla 13, Figura 15*). Brevemente, MZF1 (*del inglés, Myeloid Zinc Finger 1*) es un factor de transcripción que ha sido descrito dado que puede regular la proliferación hematopoyética y la tumorigénesis. Además, tiene un rol importante en la migración, invasión y diferenciación en cáncer, regulando la expresión de una variedad de genes⁶²⁻⁶⁴. En el caso de ZKSCAN3, también pertenece a la familia de factores de transcripción del tipo dedos de zinc. Este FT ha sido descrito como un represor transcripcional, en particular como un conductor de la proliferación celular en cáncer de colon tanto *in vitro* como *in vivo* y recientemente como un represor de la autofagia^{65,66}.

ARNTL (*del inglés, Aryl hydrocarbon receptor nuclear translocator-like protein 1*) no ha sido vinculado aún con procesos tumorigénicos, pero a la fecha se conoce que es un gen relevante para el ciclo circadiano⁶⁷. Mientras, FOXP1 (*del inglés, forkhead box P1*) ha sido identificado como un posible supresor de tumores⁶⁷.

Como vemos en las *tabla 13*, varios otros sitios putativos de unión a factores de transcripción presentaron valores de matriz bastante cercanos a 1. Algunos de ellos son FOXA1/FOXA2 (SNV₁), NF-1 (SNV₂), WT1 (SNV₂₄) y MNT (SNV₃₄) (*figura 15*). En el caso de WT1 (*del inglés, Wilms Tumor 1*) se ha descrito previamente por su rol como supresor de tumores. Y el factor MNT (*Del inglés, MAX Network Transcriptional Repressor*) pertenece a la red de reguladores transcripcionales MYC/MAX/MAD y en particular actúa como un represor.

Tabla 13. Sitios putativos de unión a FTs, identificados en la región promotora estudiada de RAET1E, (Predicción realizada con MatInspector).

SNV	COORDENADA SNV (hg19)	POSICIÓN SEC. BLANCO	FACTOR DE TRANSCRIPCIÓN (FT)	LOCALIZACIÓN SITIO DE UNIÓN FT	MATRIZ DE SIMILARIDAD	SECUENCIA	HEBRA	ALELOS
1	150217232	94	FOXM1	80-96	0,999	ggcaggcAAACatgctgg	+	C/T
4	150217387	249	HMX3	232-250	0,931	ggtaaggtAAGTgggtctt	+	T/C
4	150217387	249	NKX3-2	234-252	0,983	taaggtAAGTgggtctt	+	T/C
5	150217435	297	STAT5	279-297	0,94	tactTTCTtgaagtctgt	-	A/G
6	150217507	369	KLF12	360-376	0,922	ggacaGTGGGtcaagcg	+	G/A
6	150217507	369	Nkx5-2	369-387	0,94	gctctctcAAACgcttgac	-	G/A
7	150217633	495	GATA3	495-507	0,938	gacAGATttgaga	-	T/A
8	150217657	519	NKX2-5	504-522	0,994	tgctcTGAGTggagcctgt	+	C/T
8	150217657	519	KID3	510-520	0,974	agctCCACTc	-	C/T
13	150218006	868	SPIB	862-882	0,907	cagaaaggGAAAgggggccct	+	G/T
13	150218006	868	MZF1	866-876	1	aaGGGaaagg	+	G/T
14	150218016	878	SPIB	862-882	0,907	cagaaaggGAAAgggggccct	+	G/T
14	150218016	878	C-REL	874-888	0,923	agggggcccTTCCacc	+	G/T
14	150218016	878	SOX6	876-898	0,985	gggcaACAAagggtggaagggccc	-	G/T
15	150218017	879	C-REL	874-888	0,923	agggggcccTTCCacc	+	C/T
16	150218600	1462	HMGAI	1441-1463	0,943	ttaaAATTtctgtctctttgg	-	A/T
19	150218753	1615	NR5A2	1604-1618	0,947	atgccCAAGgcccggc	+	C/A
19	150218753	1615	ZNF239	1605-1625	0,997	tcgctctcggggCCTTggca	-	C/A
20	150218887	1749	ISX	1732-1752	0,912	cccccaAATTggagagagggg	+	G/A
20	150218887	1749	CUX1	1743-1765	0,949	ggagtccCAATcgccccctctc	-	G/A
21	150219021	1883	ZNF148	1872-1894	0,948	cagcatctCCCCcaaccaccactt	+	C/T
21	150219021	1883	ZNF219	1875-1897	0,991	catctCCCCcaaccaccactt	+	C/T
21	150219021	1883	ZBED4	1877-1891	0,98	tggggttGGGGgaga	-	C/T
24	150219142	2004	ZNF148	2003-2025	0,975	atcccccaCCCCtaccctgtgc	+	T/A
24	150219142	2004	ZKSCAN3	2001-2023	1	ccatcCCCCaccctaccctgt	+	T/A
24	150219142	2004	MAZ	2003-2015	0,901	gggggtGGGGgat	-	T/A
24	150219142	2004	ZBED4	2002-2016	0,938	aggggtGGGGgatg	-	T/A
24	150219142	2004	NME1/NME2	2002-2016	0,945	agGGGtgggggatggc	-	T/A
24	150219142	2004	KLF3	2002-2016	0,973	aggGGTgggggatggc	-	T/A
24	150219142	2004	WT1	1997-2015	0,977	gggggTGGGgatggctgg	-	T/A
24	150219142	2004	KLF2	2002-2018	0,991	gtaggGGTgggggatg	-	T/A
24	150219142	2004	PLAG1	1995-2017	1	taGGGGtgggggatggctgg	-	T/A
26	150219215	2077	TFAP2A	2066-2080	0,944	gaggCCTGaggtgac	+	T/C
26	150219215	2077	SREBF1	2068-2082	0,938	cagTCACctcaggcc	-	T/C
26	150219215	2077	SPI1	2075-2095	0,962	gactctgaGAAAgcagtcacc	-	T/C
27	150219274	2136	NACA1	2132-2144	0,934	ccccAGAgagac	-	C/A
28	150219278	2140	KLF1	2137-2153	0,97	tctGGGTtggagaggc	+	G/A
28	150219278	2140	NACA1	2132-2144	0,934	ccccAGAgagac	-	G/A
29	150219281	2143	KLF1	2137-2153	0,97	tctGGGTtggagaggc	+	T/C
29	150219281	2143	NACA1	2132-2144	0,934	ccccAGAgagac	-	T/C
30	150219283	2145	KLF1	2137-2153	0,97	tctGGGTtggagaggc	+	T/C
31	150219288	2150	KLF1	2137-2153	0,97	tctGGGTtggagaggc	+	A/G
32	150219326	2188	MAZ	2186-2198	0,932	ccgggtGGGGcca	+	G/C G/A
32	150219326	2188	GRHL2	2182-2194	0,947	caAACCGgggtggg	+	G/C G/A
32	150219326	2188	SALL2	2186-2196	0,952	ccggGTGGgc	+	G/C G/A
32	150219326	2188	POU2F1	2176-2190	0,958	accATGCaaccggg	+	G/C G/A
32	150219326	2188	KLF3	2185-2201	0,99	accGGGTggggccacct	+	G/C G/A
32	150219326	2188	POU5F1	2171-2189	0,902	ccggtttGCATgggtggacg	-	G/C G/A
32	150219326	2188	GRHL1	2181-2193	0,922	ccacccGGTTtgc	-	G/C G/A
32	150219326	2188	SCX	2189-2209	0,939	cagttgacaggTGGccccacc	-	G/C G/A
32	150219326	2188	ZKSCAN3	2178-2200	1	gggtggCCCCaccggtttgcag	-	G/C G/A
33	150219339	2201	MEIS1	2200-2216	0,986	ctgtcaacTGTCacgtg	+	T/C
33	150219339	2201	MYB	2195-2215	0,988	gccacctgtCAACTgtcacgt	+	T/C
33	150219339	2201	KLF3	2185-2201	0,99	accGGTggggccacct	+	T/C
34	150219343	2205	XBP-1	2204-2224	0,903	caactgtcACGTgtgagtgt	+	A/G
34	150219343	2205	MEIS1	2200-2216	0,986	ctgtcaacTGTcagctg	+	A/G
34	150219343	2205	MYB	2195-2215	0,988	gccacctgtCAACTgtcacgt	+	A/G
34	150219343	2205	ARNTL	2205-2221	1	aaactgtcaCGTgtgagt	+	A/G
38	150219484	2346	HOXC13	2331-2347	0,911	aaattacatTAAAtcct	-	T/C
39	150219521	2383	MESP1	2378-2390	0,951	cataCATAtgttt	+	A/G
39	150219521	2383	FOXP1	2380-2396	1	caaggaaACAAtatgta	-	A/G
40	150219551	2413	WT1	2404-2422	0,926	aaaggTGGGggccccacc	-	C/T

La tabla muestra de izquierda a derecha: La variante (SNV de 1 al 42) localizada dentro de la región putativa de unión del factor de transcripción. Luego su posición en la secuencia analizada (2696 pb) y el factor de transcripción predicho en dicha región. En cada caso se describe la localización de la secuencia reconocida por el factor de transcripción. Para cada predicción se entrega el valor de la matriz de similitud construida en base a la comparación de la secuencia consenso para determinado FT versus la secuencia en la región del genoma de referencia (hg19). Los valores de esta varían entre 0 y 1. El corte se estableció para valores mayores de 0.9. En esta predicción podemos establecer si el factor de unión a la hebra forward (+) o reverse (-). Por último se indican los alelos identificados en dicha posición.

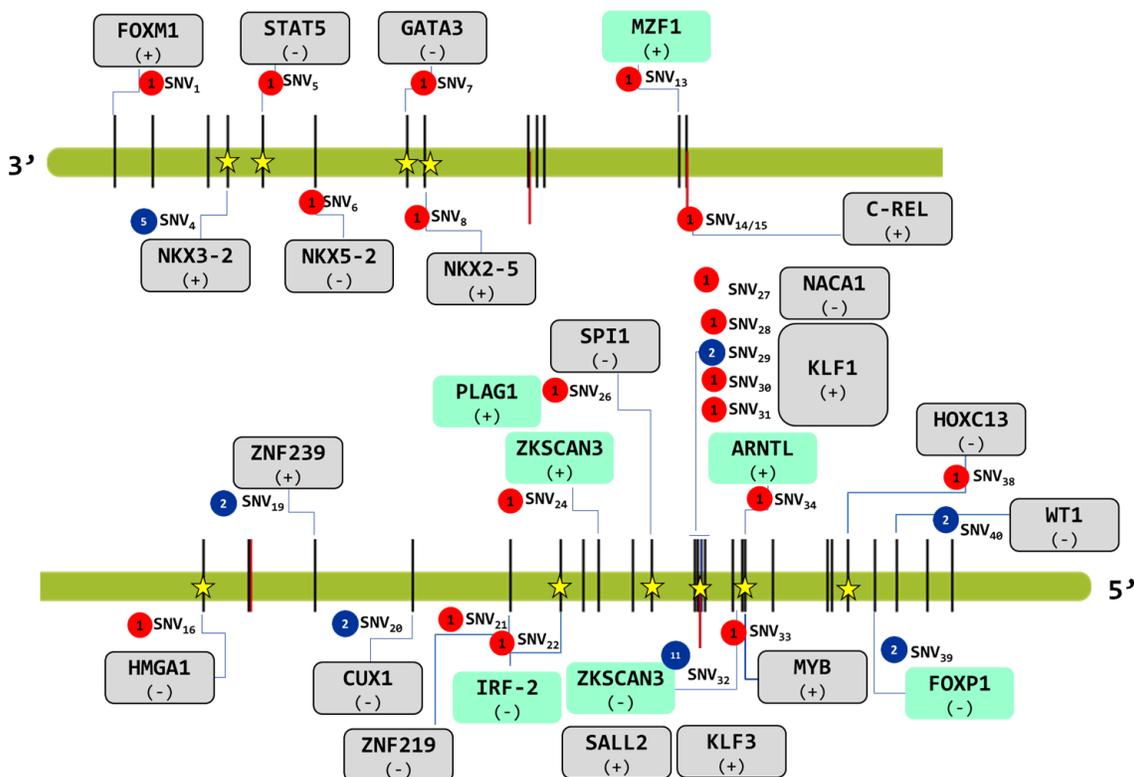


Figura 15. Predicción de sitios de unión putativos para factores de transcripción en la posición de los SNVs identificados a partir de la secuencia no mutada del promotor de RAET1E_1. El esquema anterior representa los resultados de la predicción de sitios de unión para FTs a partir de la secuencia consenso (no mutada) del promotor de RAET1E, focalizando en particular en las posiciones donde han sido identificados los SNVs. El software entrega como resultados los sitios putativos de unión para FTs, el que ha sido determinando asignando un valor de matriz de similitud. La que expresa el grado de similitud entre la secuencia blanco (nuestro promotor) y los nucleótidos de la secuencia consenso (más conservados) del sitio de unión para un FT en particular. Un valor de 1 se establece si la secuencia evaluada corresponde a los nucleótidos más conservados de cada posición de la matriz. Predicción realizada mediante PROMO 3.0 y MatInspector.

Predicción realizada mediante PROMO 3.0 y MatInspector.

3. EFECTO DE LOS SNV_s EN LOS SITIOS DE UNIÓN A FACTORES DE TRANSCRIPCIÓN.

El primer nivel de regulación de la expresión de un gen está determinado por su transcripción, proceso que está fuertemente regulado por factores de transcripción que reconocen pequeñas secuencias de ADN en su promotor. Variaciones genéticas en esta región podrían modificar los sitios de unión y por ende alterar su expresión. El objetivo consistió en evaluar el efecto de los SNVs sobre los sitios de unión putativos para factores de transcripción. Es decir, establecer mediante un análisis *in silico* si hubo generación o pérdida de sitios de unión putativos para factores de transcripción, para lo cual se utilizó el programa *Variant Analysis* de Genomatix y PROMO 3.0.

De los 42 SNVs, 23 correspondieron a las variantes ya descritas y reportadas en dbSNP, se analizaron mediante la plataforma *Variant Analysis* para lo cual utilizamos los respectivos ID. El resto de las variantes (19 no descritas) se analizó utilizando la plataforma virtual PROMO 3.0

Los resultados obtenidos mediante *Variant Analysis*, fueron filtrados seleccionando solo aquellos cuya matriz de similitud fue mayor a 0.9. Y como muestra la *figura 16* no todos los SNVs generaron cambios en los sitios de unión a factores de transcripción.

Los análisis muestran que 9 de los 42 SNVs alterarían sitios de unión putativos para factores de transcripción, lo que originaría tanto la pérdida de 6 sitios (-) como la generación de 6 nuevos sitios de unión (+). Estas variantes son los SNVs 8, 13, 14, 22, 32, 35, 36, 40 y 42 (*Figura 16*). Llamativamente observamos que algunos de estos SNVs están simultáneamente en un mismo tumor.

Entre ellos, el tumor CG07, proviene de un paciente con un estado avanzado de cáncer gástrico (T4) y con metástasis linfática en más de 15 linfonodos (N3). En particular en este tumor identificamos 5 variantes de tipo germinal, 3 de ellas únicas para este tumor. Dos de las cuales producirían la pérdida de sitios de unión para el factor de transcripción MZF1 (SNV₁₃), y NFkB (SNV₁₄). La pérdida de un sitio de unión podría asociarse a la disminución o baja expresión del ligando RAET1E, dado

que su inducción se verá alterada por la ausencia de estos sitios. A su vez en este mismo tumor identificamos una variante que genera un nuevo sitio de unión para el factor NKX 2-5 (SNV₄₂). La aparición de un nuevo sitio de unión podría, dependiendo del contexto celular, generar una inducción de RAET1E de forma crónica (en el tiempo) cuyo efecto sobre su receptor NKG2D resultaría perjudicial. Lo cual se discutirá más adelante.

Los tumores CG02 y CG45, corresponden a los tumores con el mayor número de variantes identificadas, 13 y 10 SNVs respectivamente. Además, hemos identificado un patrón de 8 SNVs (germinales) en común entre ellos, de los cuales 3 tendrían un efecto en los sitios de unión putativos para FTs (*Figura 16*). Interesantemente, ambos tumores se localizaron en la zona antral del estómago, sin embargo, el tumor CG45 es un tumor de etapa temprana (T1) que ha crecido en la capa de la submucosa y se ha diseminado a un par de ganglios linfáticos (N1), mientras que el tumor CG02 corresponde a un tumor en etapa avanzada (T4) que no se ha diseminado a los ganglios (N0).

El SNV₃₅, generaría nuevos sitios de unión para los factores AP-1 y EVI1, ambas descritas como oncoproteínas. Además, en estos tumores se generaría un nuevo sitio de unión para NF-κB (SNV₄₀), molécula que ya sabemos estaría involucrada en la inducción de alguno de los ligandos de NKG2D⁶⁸. Pero en esta misma posición podría suprimirse un sitio de unión putativo para la molécula WT-1 (SNV₄₀), conocida por su rol como supresor de tumores. Por último, el SNV₃₂ generaría la pérdida de sitios de unión para las moléculas KLF3 y Sall2, involucradas con procesos hematopoyéticos y tumorales, respectivamente. Esta variante además fue identificada en otros 9 tumores (*ver tabla 8*).

El tumor CG45 presenta además 2 SNVs distintivos (SNV₈ y SNV₃₆) también germinales, cuyo efecto sería la generación de nuevos sitios de unión para la controversial proteína p53.

El tumor CG22, es en particular un caso interesante. Es un tumor cuya localización es menos común dentro de los cánceres gástricos (Cardias), de tipo avanzado (T4) y con diseminación a varios ganglios linfáticos (N2) y un tamaño

tumoral bastante grande (mayor a 10 cm de diámetro). En este tumor identificamos 5 SNVs, 2 de ellos posiblemente de tipo somático. El SNV₂₂ mostró, tras el análisis, que esta la variación generaría la pérdida de un sitio de unión para IRF-2, un factor que participa modulando la respuesta inmune.

Por lo tanto estos resultados nos permiten concluir que efectivamente variaciones de un único nucleótido en el promotor de *RAET1E*, producirían un efecto a nivel de los sitios de unión en el ADN ya sea perdiéndose sitios de unión para FTs o bien ganando nuevos sitios.

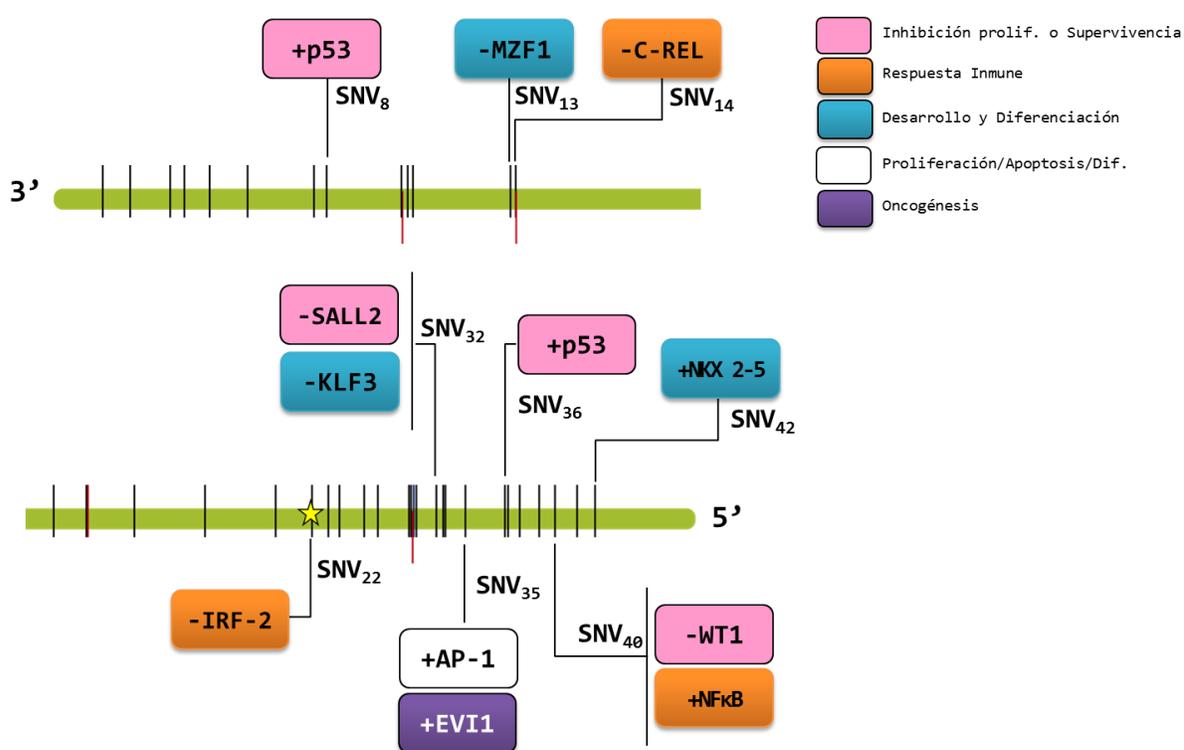


Figura 16. **Efecto de los SNVs en los sitios putativos de unión de factores de transcripción.** En verde se muestra el segmento del promotor de *RAET1E* secuenciado en cada uno de los pacientes. Cada línea sobre este segmento indica la posición de un SNV. La figura muestra los 9 SNV que originan cambios en los sitios de unión putativos, señalando en cada caso el SNV y el factor de transcripción afectado por la variación genética. El signo negativo previo al factor de transcripción indica la pérdida de sitio de unión mientras que el signo positivo indica la generación de un nuevo sitio. La estrella indica que esa variante no ha sido descrita previamente. Se han coloreado los factores de acuerdo a su participación en distintos procesos biológicos, como se indica en la leyenda. Predicción realizada mediante el programa Variant Analysis de Genomatix y PROMO 3.0.

DISCUSIÓN

Actualmente se han descrito una amplia gama de mecanismos de evasión que presentan las células tumorales para escapar al control inmunológico. Se han descrito algunos mecanismos que facilitarían la evasión del control antitumoral mediado por el sistema compuesto por el receptor de activación NKG2D y su ligandos (RAET1s/MICs). La correcta expresión de los ligandos de NKG2D en la superficie de las células tumorales es de gran relevancia para que las células efectoras, como las células NK, sean capaces de identificar tempranamente a las células tumorales para eliminarlas⁶⁹, proceso que normalmente ocurre en nuestro organismo como parte de la inmunovigilancia del cáncer.

Anteriormente se mencionó que la regulación de la expresión de los ligandos de NKG2D (NKG2DL) ocurre en varias etapas de su biogénesis²¹. La regulación de la expresión de estos ligandos (NKGDL) ocurre inicialmente a nivel transcripcional, allí una serie de proteínas conocidas como factores de transcripción se unen a la región promotora de estos ligandos para inducir su expresión en un tiempo y tejido específico. La existencia de variaciones genéticas en regiones regulatorias que alteren la transcripción ha sido establecido durante las últimas décadas y tendrían un rol en las enfermedades humanas, entre ellas el cáncer⁶¹.

En este estudio se investigó si variaciones genéticas en el promotor de *RAET1E*, un ligando del receptor NKG2D, identificadas a partir de tumores de pacientes con cáncer gástrico podrían generar cambios en su región reguladora, potenciando mecanismos de evasión de la respuesta inmune. Diversas estrategias evadirían la eliminación de las células tumorales suprimiendo la expresión de los ligandos de NKG2D. En este sentido las variaciones genéticas, como los SNVs podrían suprimir sitios de unión de factores de transcripción, y por ende disminuir la expresión de la molécula *RAET1E* en las células tumorales, alterando el reconocimiento por las células NK. Por otro lado una expresión sostenida y crónica de los ligandos llevarían a la modulación del receptor NKG2D, provocando su endocitosis y en consecuencia alterando la activación dependiente de este receptor⁷⁰.

El grupo de estudio incluyó 50 muestras de tumor de pacientes con adenocarcinoma gástrico. Un 80% de los casos presentó un cáncer en estadios avanzados en etapas T3 y T4 (*tabla 5*), lo cual coincide con lo reportado en la literatura respecto a que pese a los avances en las tecnologías de diagnóstico este tipo de cáncer es detectado después de la invasión de la *muscular propia*⁷, principalmente porque el adenocarcinoma gástrico que corresponden al 90% de los casos de cáncer gástrico progresa de manera lenta y asintomática, haciéndose evidente en etapas tardías (T3 y T4)⁷.

Sin duda alguna el avance en las tecnologías de secuenciación masiva de última generación (*NGS, del inglés Next Generation Sequencing*) permiten hoy superar limitantes biológicos como la baja pureza de una muestra tumoral, donde encontraremos además de células tumorales, células no tumorales (sanas), células estromales, vasos sanguíneos y células inmunes⁷¹, permitiendo identificar de manera robusta variantes que se encuentran en una bajo porcentaje de células dentro de esta mezcla heterogénea en la que además nos enfrentamos a más de un clon tumoral⁷². Sin embargo, aún existen varios desafíos en la detección e identificación de variantes en muestras tumorales, como las limitaciones de los algoritmos empleados, errores generados durante la secuenciación, así como los artefactos propios de la preparación de las librerías genéticas. Las enzimas más comúnmente utilizadas en la preparación de las muestras para NGS en la etapa de PCR han sido bastante caracterizadas respecto a su fidelidad y tasas de error⁷².

Diferentes estrategias han sido utilizadas para controlar los errores generados durante la secuenciación así como para identificar los distintos tipos de variaciones genéticas. En el primer caso, los artefactos propios de la técnica de secuenciación se discriminan evaluando la calidad de los datos crudos (total de lecturas generadas). En particular, los datos de secuencia entregados en la plataforma *Illumina* poseen un parámetro de calidad denominado valor de calidad *Phred* que evalúa la probabilidad del correcto llamado de bases⁷³. Este valor denominado Q_{30} indica que la probabilidad de llamar una base incorrectamente es 1 en 1000. Por lo tanto, en esta primera etapa se descartó todas las lecturas cuyos valores de calidad estuvieron bajo el Q_{30} ⁷⁴. Las siguientes etapas del análisis como el alineamiento de las lecturas al

genoma de referencia y la identificación de variaciones, dependen particularmente del tipo de muestra utilizada en el estudio, así como del diseño del mismo. Recordemos que en nuestro caso estamos frente a un estudio de re-secuenciación dirigida (*resequencing targeted*) donde sólo hemos incluido muestras de tumores. Lo que a la hora de analizar puede resultar un poco más complejo, ya que la mayoría de los flujos de trabajo de análisis utilizados emplean la información proveniente de la muestra de tumor, pareada con tejido no tumoral del mismo individuo, como sangre o tejido sano adyacente^{58,75}. De esta forma se facilita la discriminación de variantes germinales versus variaciones somáticas propias del tumor. Ya que en nuestro caso no contamos con la información de la secuenciación de tejido sano, utilizamos el algoritmo *Somatic Variant Caller* (*Illumina, Inc.*) diseñado y recomendado por Illumina para un estudio de este tipo, donde deseamos identificar variaciones con una frecuencia alélica menor al 5%.

Otro parámetro relevante a la hora de analizar los datos de secuenciación fue la profundidad de la secuenciación, que nos dice el número de veces que una región es secuenciada dentro de la misma corrida de secuenciación. En particular, se analizó la cobertura promedio para cada amplicón (*figura 13*). Es decir, las veces (X) que se leyó cada amplicón. Un total de 12 amplicones cubrieron la región total del promotor de *RAET1E* secuenciada.

Recordemos que previo al proceso de secuenciación, en la celda de flujo ocurre una clusterización o formación de grupos. Allí cada librería que incluye un amplicón (amplicón 1 al 12) se multiplica muchas veces. Esto para cada uno de los 50 pacientes. Suponiendo que en la etapa de agrupación de las librerías se mezclaron cantidades homogéneas de los 12 amplicones para cada uno de los pacientes, se espera observar un patrón mucho más homogéneo en la cobertura. Sin embargo, se observa un patrón heterogéneo, cuyas coberturas promedio van desde los 1000X a los 8000X. Al respecto, creemos que esto puede atribuirse a la etapa de preparación de librerías, donde un pool de oligos (CAT) hibrida en el ADN_g y luego todos son sometidos a las mismas condiciones de PCR, sin embargo, es probable que la eficiencia de cada par de oligos difiera, de forma que para los

amplicones 3, 6, 8,9 y 10 la eficiencia del PCR no fue la óptima. De todas formas como muestra la figura 13, se obtuvieron valores muy buenos de cobertura.

De acuerdo a lo reportado en la literatura los requerimientos de cobertura promedio en estudios de re-secuenciación dependerán del tipo de muestra que estemos analizando, del tipo de variante que nos interese identificar, así como del tamaño de la región de interés⁷⁶. En una revisión acerca de los métodos de secuenciación y el manejo bioinformático de los datos de secuenciación se recomienda para muestras de tumor una profundidad de secuenciación de al menos 125X⁵⁹. Por lo tanto para este estudio la cobertura no sería una limitante, dado que al analizar las coberturas específicas por cada amplicón y paciente se concluyó entonces que cada posición de los fragmentos cortos de ADN secuenciados (amplicones) se leyó a lo menos 200 veces.

Una vez identificadas las variantes, estas fueron sometidas a varios filtros bioinformáticos (ver tabla 4), los que descartaron variantes cuya frecuencia alélica fuese menor al 1% (Low Variant Freq), aquellas variantes identificadas en sitios con una cobertura menor a 10 lecturas (Low DP) y aquellas variantes que presentaran *Strand Bias*. Este último filtro se aplica cuando la variación identificada presenta diferencias de resultados entre las hebras (forward y reverse), lo que implica que el genotipo inferido de la hebra negativa y de la hebra positiva son significativamente diferentes⁷⁷, lo que podría llevar a la aparición de variantes falsos positivos. Una de las posibles causas sería la baja cobertura, por lo tanto es muy importante calcular previamente la cobertura estimada, de acuerdo al tipo de estudio y tipo de muestras.

Este estudio identificó 42 variantes a partir de la secuenciación de los tumores de pacientes con adenocarcinoma gástrico (*tabla 8*), de las cuales 19 no habían sido descritas previamente. De las variantes identificadas pudimos observar que 13 de ellas se presentaron en más de un tumor mientras que las 29 restantes fueron variaciones detectadas individualmente en los tumores. El hecho de que estas variantes se hayan identificado en 1 de 50 tumores, no significa que tengan menos importancia, dado que para un tumor en un paciente en particular el efecto puede ser muy significativo a nivel de la transcripción, objeto de nuestra investigación y que se discutirá más adelante.

En el contexto de la genómica y el cáncer, se han identificado en los individuos de una población variaciones germinales que podrían asociarse con una mayor o menor susceptibilidad a padecer esta enfermedad, o bien variaciones somáticas desarrolladas durante el proceso tumorigénico y que resultan relevantes tanto para el entendimiento de los mecanismos implicados en cáncer así como desde el punto de vista terapéutico⁶⁰. Para el cáncer gástrico unas pocas alteraciones germinales (gen *PLCE1*) y otras somáticas (genes *ERBB2*, *ERBB3*, *PTEN*, *PI3K/AKT/mTOR*, *FGF*, *TP53*, *CDH1* y *MET*) han emergido y sólo algunas de ellas han sido objeto de interés clínico⁶⁰.

Una vez identificadas las variantes en el promotor de *RAET1E*, se analizó su frecuencia alélica. De forma complementaria se utilizó la base de datos dbSNP y COSMIC, determinando que 25 variantes eran germinales y 17 somáticas (Tablas 9 y 10).

En cuanto a las variantes germinales, fue evidente la presencia de ciertos bloques de SNVs. Así por ejemplo un bloque de 3 variantes germinales (3, 11 y 42, tabla 8) que se detectó en 36 de 50 pacientes. Sospechamos que podría tratarse de un haplotipo, el que a su vez podría ser un haplotipo de riesgo para la población estudiada, para lo cual se requiere realizar mayores análisis. En particular, el análisis del efecto de los SNVs sobre los sitios putativos de unión de FTs mostró que el *SNV*₄₂, origina un sitio de unión putativo para NKX 2-5, un factor involucrado en procesos de diferenciación y desarrollo⁷⁸.

El segundo bloque identificado, corresponde a un grupo de 8 SNVs (todos germinales), que llamativamente se identificó en 2 de los 50 pacientes (CG02 y CG45), y 7 de estas variantes (SNVs 19, 20, 23, 29, 35, 39 y 40) no han sido identificadas en población asiática. Población con la cual compartimos altos índices de mortalidad por cáncer gástrico. Por lo tanto, es posible que estas variaciones resulten importantes dentro de nuestra población. Más tarde cuando se *analizó in silico* el efecto de estas variantes, 3 de ellas comunes para estos dos tumores (*SNV*₃₂, *SNV*₃₅ y *SNV*₄₀) modificaron sitios de unión para FTs. Tanto WT1 como Sall2, involucradas en la inhibición de la proliferación y sobrevivencia, pierden sitios de reconocimiento producto de los SNVs 32 y 40, pero llamativamente el *SNV*₄₀

generaría un sitio de unión para NF-KB. Molécula que ya ha sido descrita como un regulador transcripcional de los ligandos de NKG2D, sin embargo el efecto del sitio adicional generado, debería ser evaluado *in vitro*. Adicionalmente el SNV₃₅ genera sitios nuevos de unión para AP-1 y EVI1, ambas proteínas descritas como oncoproteínas, las que inducirían la expresión del ligando RAET1E. Llamativamente, el tumor CG45, que corresponde a un tumor primario, posee dos SNVs distintivos (8 y 36, respectivamente) que generan la aparición de nuevos sitios para el supresor de tumores p53. A nivel clínico ambos tumores se localizaron en el antro, sin embargo mientras el tumor CG45 posee una clasificación T1 el tumor CG02 está clasificado como T4. Claramente vemos que dentro de cada tumor existe una heterogeneidad desde el punto de vista genético lo cual puede verse ejemplificada en estos dos casos.

Dentro del grupo de variantes somáticas, la variante SNV₂₂, identificada en un único paciente (CG22) originó la supresión de un sitio de unión para IRF-2, involucrado en la respuesta inmune.

Desde el punto de vista inmunológico, es pertinente discutir cómo estos cambios en los sitios putativos de unión para FT podrían vincularse a la respuesta inmune y por ende favorecer al desarrollo tumoral.

Llamativamente, cuando analizamos la presencia de estas 9 variantes en el grupo de tumores estudiado, pudimos identificar en algunos pacientes la presencia simultánea de más de una de estos 9 SNVs.

El tumor CG07 porta 5 SNVs de tipo germinal en el promotor RAET1_1, 3 de los cuales alterarían sitios de unión para factores de transcripción. Uno de ellos corresponde a C-Rel, miembro de la familia factores NF-kB (REF), relevante en la respuesta inmune. En este caso el SNV₁₄ suprime este sitio. Por lo tanto al no unirse al promotor no induciría la transcripción de RAET1E, probablemente disminuyendo su expresión. Si pensamos esto en un contexto de tumorigénesis, donde necesitamos inducir la expresión de esta molécula, el efecto neto sería una disminución del reconocimiento de las células NK y el escape del control tumoral, favoreciendo el desarrollo tumoral. Las dos otras variantes generan nuevos sitios de

unión para FT para los cuales no se tienen antecedentes por literatura de su vinculación con la regulación de alguno de los ligandos de NKG2D.

Por otro lado, los tumores CG02 y CG45, en los cuales identificamos la presencia de un patrón de 8 SNVs de tipo germinal en común, y 3 de ellos estarían alterando sitios de unión para FT. Los SNV₃₂ y SNV₄₀ generarían la ausencia de un sitio de unión putativo para Sall2 y WT-1, ambos identificados como supresores de tumor. Si bien ninguno de estos dos factores se ha identificado antes como un regulador del resto de los ligandos de NKG2D, por literatura sabemos que la molécula p53, induciría directa o indirectamente su transcripción²⁷. Por lo tanto creemos que la ausencia de estos sitios repercutiría en una disminución de su transcripción y por lo tanto en su expresión en superficie. Esto a su vez también favorecería la evasión inmune, al disminuir la activación de las células NK mediada por la unión de RAET1E y NKG2D. En el caso del tumor CG45, este mostró la generación de 2 nuevos sitios de unión para el FT p53. Desde el punto de vista inmunológico, el aumento de la expresión de RAET1E en el tiempo también corresponde a un mecanismo de evasión tumoral. Se ha descrito que la expresión prolongada y crónica de los ligandos de NKG2D promovería la endocitosis del receptor, regulando negativamente su expresión en la superficie de las células citolíticas afectando la función de las células NK, promoviendo finalmente el escape de las células tumorales.

Interesantemente de todos los casos analizados, en un tumor (CG22), identificamos una variante somática, el SNV₂₂, cuyo efecto fue la supresión del sitio de unión putativo para IRF-2, un factor de transcripción involucrado en la respuesta inmune. En este caso nos llama la atención de que no se identificaron muchas variaciones para este tumor, que se encuentra en un estadio avanzado T4N2M0. Por lo cual creemos que la ausencia de este factor producto de la variación identificada es relevante para la evasión del control inmune.

En nuestros análisis pudimos evidenciar que no habría correlación en cuanto al número de SNVs y el estadio del tumor. Más bien lo que se observa es que la progresión del cáncer gástrico dependería del tipo de factores de transcripción alterados por las variaciones genéticas. Es decir, que la ganancia o pérdida de

ciertos factores generaría mecanismos evasivos más potentes que otros. Y aun cuando estos datos deben ser evaluados *in vitro*, podríamos inferir que lo que prima es la combinación de factores afectados.

Entonces hasta aquí hemos visto que la presencia de los alelos alternativos producto de la variante identificada (SNV) generaría en la mayoría de los casos la pérdida de sitios de unión a factores de transcripción, mientras que en otros casos se habría generado la ganancia de sitios de unión. En consecuencia, creemos que esto podría generar cambios en la regulación del gen *RAET1E*, al ser inducido por algunos de estos factores de transcripción alterando la respuesta inmune innata mediada por este ligando en los pacientes con adenocarcinoma gástrico.

A raíz de estudio esperamos que investigaciones posteriores puedan identificar *in vitro* aquellas variantes que modifiquen la actividad promotora de *RAET1E*, además de estudios que permitan comprobar los sitios de unión para factores de transcripción identificados mediante el análisis computacional que hemos realizado.

CONCLUSIONES

- Esta investigación identificó 42 variaciones en el promotor de *RAET1E* en tumores de pacientes con cáncer gástrico.

- Los análisis *in silico* determinaron que 9 de las 42 variantes genéticas identificadas en el promotor de *RAET1E_1* podrían causar tanto la pérdida como la generación de sitios de unión a factores de transcripción, tales como:
 1. Supresores de tumor (Sall2, WT1 y p53)
 2. Factores involucrados en la respuesta inflamatoria (NFκB/IRF-2/c-Rel)
 3. Oncoproteínas (EVI1)
 4. Factores relevantes para procesos de desarrollo y diferenciación (KLF3 y NKX 2-5)
 5. Reguladores de la apoptosis, proliferación y diferenciación (AP-1)

- La supresión de sitios de unión para determinados factores de transcripción puede disminuir la transcripción del ligando RAET1E, disminuyendo su expresión en la superficie de las células tumorales afectando su reconocimiento por las células citolíticas como las células NK. En consecuencia habría una disminución de la activación de la citotoxicidad de estas células mediadas por la unión del ligando RAET1E y su receptor NKG2D, favoreciendo el escape de las células tumorales al control inmunológico.

- La generación de nuevos sitios de unión para factores de transcripción, promovería la inducción de su transcripción, posiblemente aumentando su expresión en la superficie de las células tumorales. La expresión sostenida y prolongada de RAET1E, regularía negativamente la expresión del receptor NKG2D en las células citolíticas, como la célula NK, comprometiendo su citotoxicidad.

ANEXOS

- A. Frecuencias Alélicas de las variantes: SNV₃, SNV₁₁, SNV₄₂ y SNV₃₂.
- B. Acta de aprobación y Consentimiento Informado Pacientes con Diagnóstico de Cáncer Gástrico, del Comité de Ética en Investigación en Seres Humanos de la Facultad de medicina de la Universidad de Chile.
- C. Resumen Congreso generado por la Tesis.

ANEXO A. FRECUENCIAS ALÉLICAS DE LAS VARIANTES: SNV3, SNV11, SNV42 Y SNV32.

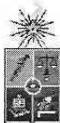
N° SNV	Muestra	Genotipo	Variante	Frecuencia Alélica	ID dbSNP	Frecuencias Poblacionales			
						F.A Amr	F.A Asn	F.A Af	A.F Eur
3	CG02	HETEROIGOTO	C/T	39,90	rs9371544	50	85	33	41
	CG08			48,83					
	CG10			50,23					
	CG13			48,83					
	CG15			75,82					
	CG18			51,09					
	CG19			52,50					
	CG21			56,39					
	CG22			48,20					
	CG23			46,99					
	CG24			46,39					
	CG28			48,18					
	CG30			50,30					
	CG32			48,27					
	CG37			36,43					
	CG40			62,45					
	CG50			52,09					
	CG01	HOMOCIGOTO	C/C	99,82					
	CG06			99,70					
	CG07			99,89					
	CG09			99,75					
	CG14			99,75					
	CG16			99,78					
	CG17			99,78					
	CG33			99,61					
	CG34			99,71					
	CG35			99,86					
	CG38			99,76					
	CG41			99,68					
	CG42			99,79					
	CG43			99,83					
	CG44			99,63					
	CG49			99,83					
CG51	99,78								
CG52	99,81								
CG55	99,88								
CG56	99,79								
CG57	99,95								
CG59	99,77								
CG60	99,84								

N° SNV	Muestra	Genotipo	Variante	Frecuencia Alélica	ID dbSNP	Frecuencias Poblacionales			
						F.A Amr	F.A Asn	F.A Af	A.F Eur
11	CG02	HETEROCIGOTO	T/C	40,82	rs6935829	49	85	33	41
	CG08			48,58					
	CG10			49,28					
	CG13			50,86					
	CG15			75,97					
	CG18			49,83					
	CG19			55,57					
	CG21			67,74					
	CG22			52,85					
	CG23			50,82					
	CG24			48,77					
	CG28			51,65					
	CG30			48,19					
	CG32			49,95					
	CG37			39,7					
	CG40			60,64					
	CG50			48,18					
	CG52			HOMOCIGOTO					
	CG01	99,5							
	CG06	99,53							
	CG07	99,4							
	CG09	99,49							
	CG14	99,52							
	CG16	99,57							
	CG17	99,57							
	CG33	99,63							
	CG34	99,33							
	CG35	99,5							
	CG38	99,51							
	CG41	99,45							
	CG42	99,34							
	CG43	99,11							
	CG44	99,26							
	CG49	99,35							
	CG51	99,56							
	CG55	99,47							
CG56	99,34								
CG57	99,5								
CG59	99,41								
CG60	99,48								

N° SNV	Muestra	Genotipo	Variante	Frecuencia Alélica	ID dbSNP	Frecuencias Poblacionales			
						F.A Amr	F.A Asn	F.A Af	A.F Eur
42	CG02	het	G>G/A	23,66	rs12528243	37	32	14	26
	CG06	het	G>G/A	51,62					
	CG07	het	G>G/A	49,33					
	CG08	het	G>G/A	52,22					
	CG13	het	G>G/A	48,52					
	CG14	het	G>G/A	54,94					
	CG18	het	G>G/A	47,80					
	CG21	het	G>G/A	72,69					
	CG22	het	G>G/A	51,40					
	CG23	het	G>G/A	51,29					
	CG24	het	G>G/A	49,33					
	CG28	het	G>G/A	50,16					
	CG30	het	G>G/A	46,41					
	CG37	het	G>G/A	40,49					
	CG40	het	G>G/A	61,81					
	CG44	het	G>G/A	49,38					
	CG50	het	G>G/A	46,40					
	CG51	het	G>G/A	52,97					
	CG52	het	G>G/A	50,38					
	CG55	het	G>G/A	42,83					
	CG56	het	G>G/A	49,84					
	CG57	het	G>G/A	55,86					
	CG01	hom	G>A/A	99,19					
	CG09	hom	G>A/A	99,51					
	CG10	hom	G>A/A	99,23					
	CG16	hom	G>A/A	99,64					
	CG17	hom	G>A/A	99,64					
	CG33	hom	G>A/A	99,64					
	CG35	hom	G>A/A	99,45					
	CG38	hom	G>A/A	99,17					
CG41	hom	G>A/A	99,72						
CG42	hom	G>A/A	99,17						
CG43	hom	G>A/A	98,77						
CG49	hom	G>A/A	99,27						
CG59	hom	G>A/A	99,15						
CG60	hom	G>A/A	99,60						

N° SNV	Muestra	Genotipo	Variante	Frecuencia Alélica	ID dbSNP	Frecuencias Poblacionales			
						F.A Amr	F.A Asn	F.A Af	A.F Eur
32	CG02	HETEROCIGOTOS	G/C	53,68	rs7761734	0	0	0	0
	CG14		G/A	40,50		13	53	7	14
	CG15			76,39					
	CG19			60,39					
	CG32			48,77					
	CG34			52,91					
	CG44			56,87					
	CG45			48,16					
	CG51			45,75					
	CG55			47,81					
	CG56			51,53					

**ANEXO B. ACTA APROBACIÓN Y CONSENTIMIENTO INFORMADO PACIENTES
DIAGNOSTICADOS CON CG, DEL COMITÉ ÉTICA EN INVESTIGACIÓN EN
SERES HUMANOS, FACULTAD MEDICINA, UNIVERSIDAD DE CHILE.**



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



31 MAYO 2011

1/2

**ACTA DE APROBACIÓN
PROYECTO DE INVESTIGACIÓN SERES HUMANOS**

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

Dr. Manuel Oyarzún G., Médico Neumólogo, Presidente
Sra. Marianne Gaudlitz H., Enfermera, Mg. Humanidades, Vicepresidente
Dr. Hugo Amigo C., Ph. D., Especialista en Salud Pública
Dr. Leandro 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

2/2



31 MAYO 2011

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

- a) Carácter de la población estudiada: Investigación no terapéutica, población no cautiva.
- b) Utilidad del Proyecto: 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



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 mi 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 molestias tales como dolor, hemorragia, o reacción alérgica por el antiséptico utilizado. Sin embargo, la muestra será tomada por una Enfermera Universitaria con experiencia en este tipo de procedimientos, quien tomará todas las precauciones para minimizar las posibilidades que esto ocurra.

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

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

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

31 MAYO 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 C. RESUMEN CONGRESO GENERADO POR LA TESIS.

Genetic variation in the promoter *RAET1E* promoter identified in patients with gastric adenocarcinoma may alter binding sites of transcription factors.

Marcela Morales Gedda¹, Valentina Gárate Calderón¹, Jessica Toro Carrasco², Carolina Hager Ribeiro¹, Patricio González Hormazabal³, Ricardo Armisen Yañez², María Carmen Molina Sampayo¹.

¹Laboratorio de Inmunovigilancia y Evasión Inmune, Programa Disciplinario de Inmunología, ICBM, Facultad de Medicina, Universidad De Chile.²Laboratorio de Patología Molecular y Genómica del Cáncer, Centro de Investigación y Tratamiento del Cáncer, ICBM, Facultad de Medicina, Universidad De Chile.³Programa de Genética Humana, ICBM, Facultad de Medicina, Universidad De Chile.
***marcemoralesg@gmail.com**

In Chile, stomach cancer is the leading cause of death due to malignant tumors, with around 3,000 people dying each year as a consequence of this disease. The immune system can recognize and eliminate transformed cells in a process named immunosurveillance. Natural Killer (NK) cells are very important in this process. They present the activating receptor NKG2D, which detects self-molecules induced in conditions of target cell stress, such as malignant transformation. In humans, there are 8 ligands to NKG2D: MIC and RAET1 molecules. Several polymorphisms have been described for these ligands, some of which have been associated with an increased risk to develop cancer and could have implications with potential mechanisms of immune evasion. Genetic variations in their promoter regions could alter expression of these ligands in a cancer context. To identify single nucleotide variations (SNV) in promoter region of NKG2D ligands that may generate changes in transcription factor binding sites in gastric cancer, we analyzed the promoter region of RAET1E in 50 tumor samples of gastric adenocarcinoma through next-generation sequencing (NGS). Further, we evaluated the effect of this variation by in silico analysis. We identified 42 SNVs, some of which have not been previously reported. Several variations identified in this study may cause either the loss or the generation of transcription factor binding sites, such as p53 and NFκB, both involved in the cellular response to stress stimuli. Thus, our investigation provide evidence for the relevance of polymorphism in the regulatory regions of RAET1E gene in gastric cancer

Key words: gastric cancer, RAET1E, SNV, NGS and polymorphism.

BIBLIOGRAFÍA

1. Stewart, B. W. & Wild, C. P. *World Cancer Report 2014. World Health Organization: Geneva* (2014). doi:9283204298
2. Thrumurthy, S. G. & Chaudry, M. A. The diagnosis and management of gastric cancer. *BMJ* **6367**, 1–9 (2013).
3. Torre, L. A. *et al.* Global cancer statistics, 2012. *CA. Cancer J. Clin.* **65**, 87–108 (2015).
4. Ferlay, J. *et al.* GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase. No. 11 [Internet]. Lyon, France: *International Agency for Research on Cancer*. **11**, <http://globocan.iarc.f> (2013).
5. Minsal. Guía Clínica: Cáncer Gástrico. *Ser. Guías Clínicas MINSAL* 2–53 (2010).
6. Alfonso Calvo, B. Diagnóstico precoz del cáncer gástrico estrategias de prevención secundaria y dificultades del diagnóstico de lesiones precoces. *Rev. Médica Clínica Las Condes* **22**, 477–484 (2011).
7. Nagini, S. Carcinoma of the stomach: A review of epidemiology, pathogenesis, molecular genetics and chemoprevention. *World J. Gastrointest. Oncol.* **4**, 156–169 (2012).
8. Médica, S. E. de O. Cáncer Gástrico. (2015). at <<http://www.seom.org/en/informacion-sobre-el-cancer/info-tipos-cancer/digestivo/estomago?start=6#content>>
9. Dicken, B. J. *et al.* Gastric Adenocarcinoma. Review and Considerations for Future Directions. *Ann. Surg.* **241**, 27–39 (2005).
10. Takahashi, T., Saikawa, Y. & Kitagawa, Y. Gastric cancer: current status of diagnosis and treatment. *Cancers (Basel)*. 48–63 (2013). doi:10.3390/cancers5010048
11. Cavallo, F., De Giovanni, C., Nanni, P., Forni, G. & Lollini, P.-L. 2011: the Immune Hallmarks of Cancer. *Cancer Immunol. Immunother.* **60**, 319–326 (2011).
12. López-Soto, A., Huergo-Zapico, L., Acebes-Huerta, A., Villa-Alvarez, M. & Gonzalez, S. NKG2D signaling in cancer immunosurveillance. *Int. J. Cancer* **136**, 1741–1750 (2015).
13. Chitadze, G., Bhat, J., Lettau, M., Janssen, O. & Kabelitz, D. Generation of Soluble NKG2D Ligands: Proteolytic Cleavage, Exosome Secretion and Functional Implications. *Scand. J. Immunol.* **78**, 120–129 (2013).
14. Waldhauer, I. & Steinle, A. NK cells and cancer immunosurveillance. *Oncogene* **27**, 5932–5943 (2008).
15. Cerboni, C. *et al.* The DNA damage response: A common pathway in the regulation of NKG2D and DNAM-1 ligand expression in normal, infected, and cancer cells. *Front. Immunol.* **4**, 508 (2013).
16. Guerra, N. *et al.* NKG2D-deficient mice are defective in tumor surveillance in models of spontaneous malignancy. *Immunity* **28**, 571–80 (2008).
17. Saito, H., Osaki, T. & Ikeguchi, M. Decreased NKG2D expression on NK cells correlates with impaired NK cell function in patients with gastric cancer. *Gastric Cancer* **15**, 27–33 (2012).
18. Osaki, T. *et al.* Decreased NKG2D expression on CD8+ T cell is involved in immune evasion in patients with gastric cancer. *Clin. Cancer Res.* **13**, 382–7 (2007).
19. Fernández-Messina, L., Reyburn, H. T. & Valés-Gómez, M. Human NKG2D-ligands: cell biology strategies to ensure immune recognition. *Front. Immunol.* **3**, 1–9 (2012).
20. Carapito, R. & Bahram, S. Genetics, genomics, and evolutionary biology of NKG2D ligands. *Immunol. Rev.* **267**, 88–116 (2015).

21. Raulet, D. H., Gasser, S., Gowen, B. G., Deng, W. & Jung, H. *Regulation of ligands for the NKG2D activating receptor. Annual review of immunology* **31**, (2013).
22. Wu, J. & Wang, X. in *Current Cancer Treatment - Novel Beyond Conventional Approaches* (ed. Oner, O.) 810 (InTech, 2011). at <<http://www.intechopen.com/books/current-cancer-treatment-novel-beyond-conventional-approaches/nkg2d-based-cancer-immunotherapy>>
23. Nowbakht, P. *et al.* Ligands for natural killer cell activating receptors are expressed upon maturation of normal myelomonocytic cells but are low in acute myeloid leukemias. *Blo* **105**, 3615–3622 (2005).
24. Poggi, a. *et al.* Interaction between Human NK Cells and Bone Marrow Stromal Cells Induces NK Cell Triggering: Role of NKp30 and NKG2D Receptors. *J. Immunol.* **175**, 6352–6360 (2005).
25. Gasser, S., Orsulic, S., Brown, E. J. & Raulet, D. H. The DNA damage pathway regulates innate immune system ligands of the NKG2D receptor. *Nature* **436**, 1186–90 (2005).
26. Chan, C. J., Smyth, M. J. & Martinet, L. Molecular mechanisms of natural killer cell activation in response to cellular stress. *Cell Death Differ.* **21**, 5–14 (2014).
27. Iannello, A. & Raulet, D. H. Immune Surveillance of Unhealthy Cells by Natural Killer Cells. *Cold Spring Harb. Symp. Quant. Biol.* (2013). doi:10.1101/sqb.2013.78.020255
28. Venkataraman, G. M., Suci, D., Groh, V., Boss, J. M. & Spies, T. Promoter region architecture and transcriptional regulation of the genes for the MHC class I-related chain A and B ligands of NKG2D. *J. Immunol.* **178**, 961–9 (2007).
29. Iannello, A. & Raulet, D. H. Immune Surveillance of Unhealthy Cells by Natural Killer Cells. *Cold Spring Harb. Symp. Quant. Biol.* (2013). doi:10.1101/sqb.2013.78.020255
30. Heinemann, A. *et al.* Tumor suppressive microRNAs miR-34a/c control cancer cell expression of ULBP2, a stress-induced ligand of the natural killer cell receptor NKG2D. *Cancer Res.* **72**, 460–71 (2012).
31. López-Soto, A., Quiñones-Lombraña, A., López-Arbesú, R., López-Larrea, C. & González, S. Transcriptional regulation of ULBP1, a human ligand of the NKG2D receptor. *J. Biol. Chem.* **281**, 30419–30 (2006).
32. Hamerman, J. a, Ogasawara, K. & Lanier, L. L. Cutting edge: Toll-like receptor signaling in macrophages induces ligands for the NKG2D receptor. *J. Immunol.* **172**, 2001–5 (2004).
33. Eisele, G. *et al.* TGF-beta and metalloproteinases differentially suppress NKG2D ligand surface expression on malignant glioma cells. *Brain* **129**, 2416–25 (2006).
34. Serrano, A. E. *et al.* Interleukin 10 decreases MICA expression on melanoma cell surface. *Immunol. Cell Biol.* **89**, 447–57 (2011).
35. Boutet, P. *et al.* Cutting edge: the metalloproteinase ADAM17/TNF-alpha-converting enzyme regulates proteolytic shedding of the MHC class I-related chain B protein. *J. Immunol.* **182**, 49–53 (2009).
36. Stern-Ginossar, N. *et al.* Human microRNAs regulate stress-induced immune responses mediated by the receptor NKG2D. *Nat. Immunol.* **9**, 1065–73 (2008).
37. Choy, M. K. & Phipps, M. E. MICA polymorphism: biology and importance in immunity and disease. *Trends Mol. Med.* **16**, 97–106 (2010).
38. Douik, H. *et al.* Association of MICA-129 polymorphism with nasopharyngeal cancer risk in a Tunisian population. *Hum. Immunol.* **70**, 45–8 (2009).
39. Rodríguez-Rodero, S. *et al.* Transcriptional regulation of MICA and MICB: a novel polymorphism in MICB promoter alters transcriptional regulation by Sp1. *Eur. J. Immunol.* **37**, 1938–53 (2007).

40. Cox, S. T. *et al.* RAET1/ULBP alleles and haplotypes among Kolla South American Indians. *Hum. Immunol.* **74**, 775–82 (2013).
41. Romphruk, A. V. *et al.* Polymorphisms of NKG2D ligands: diverse RAET1/ULBP genes in Northeastern Thais. *Immunogenetics* **61**, 611–617 (2009).
42. Antoun, A. *et al.* 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. *Hum. Immunol.* **71**, 610–620 (2010).
43. Cox, S. T., Madrigal, J. a & Saudemont, a. Three novel allelic variants of the RAET1E/ULBP4 gene in humans. *Tissue Antigens* **80**, 390–2 (2012).
44. Eagle, R. a, Traherne, J. a, Ashiru, O., Wills, M. R. & Trowsdale, J. Regulation of NKG2D ligand gene expression. *Hum. Immunol.* **67**, 159–69 (2006).
45. Cox, S. T. *et al.* Characterization of 5' promoter and exon 1-3 polymorphism of the RAET1E gene. *Hum. Immunol.* **77**, 96–103 (2015).
46. Geer, L. Y. *et al.* The NCBI BioSystems database. *Nucleic Acids Res.* **38**, D492–D496 (2010).
47. Cao, W. *et al.* RAET1E2, a Soluble Isoform of the UL16-binding Protein RAET1E Produced by Tumor Cells, Inhibits NKG2D-mediated NK Cytotoxicity. *J. Biol. Chem.* **282**, 18922–18928 (2007).
48. Chalupny, N. J., Sutherland, C. L., Lawrence, W. a., Rein-Weston, A. & Cosman, D. ULBP4 is a novel ligand for human NKG2D. *Biochem. Biophys. Res. Commun.* **305**, 129–135 (2003).
49. McGilvray, R. W. *et al.* ULBP2 and RAET1E NKG2D ligands are independent predictors of poor prognosis in ovarian cancer patients. *Int. J. Cancer* **127**, 1412–1420 (2010).
50. Dreos, R., Ambrosini, G., Perier, R. C. & Bucher, P. The Eukaryotic Promoter Database: expansion of EPDnew and new promoter analysis tools. *Nucleic Acids Res.* **43**, D92–D96 (2014).
51. Hon, G. C., Hawkins, R. D. & Ren, B. Predictive chromatin signatures in the mammalian genome. *Hum. Mol. Genet.* **18**, R195–201 (2009).
52. Kent, W. J. *et al.* The Human Genome Browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
53. Cock, P. J. A., Fields, C. J., Goto, N., Heuer, M. L. & Rice, P. M. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res.* **38**, 1767–1771 (2010).
54. Illumina. Somatic Variant Caller. **1**, 2 (2014).
55. Messeguer, X. & Escudero, R. PROMO: detection of known transcription regulatory elements using species-tailored. *Bioinforma. Appl. Note* **18**, 333–334 (2002).
56. Wingender, E. *et al.* The TRANSFAC system on gene expression regulation. *Nucleic Acids Res.* **29**, 281–283 (2001).
57. Cartharius, K. *et al.* MatInspector and beyond: Promoter analysis based on transcription factor binding sites. *Bioinformatics* **21**, 2933–2942 (2005).
58. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
59. Bao, R. *et al.* Review of Current Methods, Applications, and Data Management for the Bioinformatics Analysis of Whole Exome Sequencing. *Lib. Acad.* **13**, 67–82 (2014).
60. Wadhwa, R. *et al.* Gastric cancer-molecular and clinical dimensions. *Nat. Rev. Clin. Oncol.* **10**, 643–55 (2013).
61. Carlton, V. E. H., Ireland, J. S., Useche, F. & Faham, M. Functional single nucleotide polymorphism-based association studies. *Hum. Genomics* **2**, 391–402 (2006).

62. Gaboli, M. *et al.* Mzf1 controls cell proliferation and tumorigenesis. *Genes Dev.* **15**, 1625–1630 (2001).
63. Deng, Y. *et al.* p53/PIK transcriptionally activated by MZF1 promotes colorectal cancer cell proliferation. *Biomed Res Int* **2013**, 868131 (2013).
64. Vishwamitra, D. *et al.* The transcription factors IκB-1 and MZF1 downregulate IGF-1R expression in NPM-ALK+ T-cell lymphoma. *Mol. Cancer* **14**, 1–17 (2015).
65. Chauhan, S. *et al.* ZKSCAN3 is a master transcriptional repressor of autophagy. *Mol. Cell* **50**, 16–28 (2013).
66. Zhang, X. *et al.* The zinc finger transcription factor ZKSCAN3 promotes prostate cancer cell migration. *Int. J. Biochem. Cell Biol.* **44**, 1166–73 (2012).
67. Rebhan, M., Chalifa-Caspi, V., Prilusky, J. & Lancet, D. GeneCards: integrating information about genes, proteins and diseases. *Trends Genet.* **13**, 163 (1997).
68. Raulet, D. H., Gasser, S., Gowen, B. G., Deng, W. & Jung, H. *Regulation of ligands for the NKG2D activating receptor. Annual review of immunology* **31**, (2013).
69. González, S., López-Soto, A., Suarez-Alvarez, B., López-Vázquez, A. & López-Larrea, C. NKG2D ligands: key targets of the immune response. *Trends Immunol.* **29**, 397–403 (2008).
70. Zhang, J., Basher, F. & Wu, J. D. NKG2D Ligands in Tumor Immunity: Two Sides of a Coin. *Front. Immunol.* **6**, 1–7 (2015).
71. Mardis, E. R. Genome sequencing and cancer. *Curr. Opin. Genet. Dev.* **22**, 245–250 (2012).
72. Costello, M. *et al.* Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. **41**, 1–12 (2013).
73. Pabinger, S. *et al.* A survey of tools for variant analysis of next-generation genome sequencing data. *Brief. Bioinform.* **15**, 256–278 (2014).
74. Rutvisuttinunt, W. *et al.* Simultaneous and complete genome sequencing of influenza A and B with high coverage by Illumina MiSeq Platform. *J. Virol. Methods* **193**, 394–404 (2013).
75. Roberts, N. D. *et al.* A comparative analysis of algorithms for somatic SNV detection in cancer. *Bioinformatics* **29**, 2223–2230 (2013).
76. Sims, D., Sudbery, I., Iltis, N. E., Heger, A. & Ponting, C. P. Sequencing depth and coverage: key considerations in genomic analyses. *Nat. Rev. Genet.* **15**, 121–32 (2014).
77. Guo, Y. *et al.* The effect of strand bias in Illumina short-read sequencing data. *BMC Genomics* **13**, 666 (2012).
78. Zhou, B., Giese, A. von, Ma, Q., Rivera-Feliciano, J. & Pu, W. T. Nkx2-5- and Isl1-expressing cardiac progenitors contribute to proepicardium. *Biochem. Biophys. Res. Commun.* **375**, 450–453 (2008).