



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Escola Tècnica Superior
d'Enginyeria Agronòmica i del Medi Natural

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería Agronómica
y del Medio Natural

Análisis de Alteraciones de Splicing en Genes de
Predisposición Germinal a Neoplasia Mieloide

Trabajo Fin de Grado

Grado en Biotecnología

AUTOR/A: López Nieto, Pablo

Tutor/a: Giraldo Reboloso, Esther

Cotutor/a externo: Liquori, Alessandro

Director/a Experimental: Martínez Valiente, Cristina

CURSO ACADÉMICO: 2023/2024



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Instituto de Investigación
Sanitaria La Fe

Universitat Politècnica de València

Escuela Técnica Superior de Ingeniería Agronómica y del Medio Natural

Grado en Biotecnología

Trabajo de Fin de Grado

**ANÁLISIS DE ALTERACIONES DE *SPLICING* EN GENES DE
PREDISPOSICIÓN GERMINAL A NEOPLASIA MIELOIDE**

Curso académico 2023-2024

Autor: Pablo López Nieto

Tutora académica: Dra. Esther Giraldo Reboloso

Tutor externo: Dr. Alessandro Liquori

Directora experimental: Dra. Cristina Martínez Valiente

TÍTULO

Análisis de Alteraciones de *Splicing* en Genes de Predisposición Germinal a Neoplasia Mieloide

Analysis of Alternative Splicing Alterations in Germline Predisposition Genes to Myeloid Neoplasms

Anàlisi d'Alteracions de *Splicing* en Gens de Predisposició Germinal a Neoplàsia Mieloide

RESUMEN

Las neoplasias mieloides (NM) con predisposición germinal se diagnostican cada vez con más frecuencia, constituyendo aproximadamente el 10% de los casos de síndrome mielodisplásico (SMD) y leucemia mieloide aguda (LMA). Las mutaciones germinales en genes como *GATA2*, *ETV6*, *DDX41*, *SAMD9* y *SAMD9L* están implicadas en esta predisposición. La deficiencia de *GATA2*, un síndrome genético raro resultado de mutaciones en el gen *GATA2*, impacta significativamente en el desarrollo y función de las células sanguíneas y del sistema inmunológico. Caracterizada por inmunodeficiencia, fallo de médula ósea, SMD, enfermedad pulmonar, linfedema, y una mayor susceptibilidad a infecciones y cánceres, la deficiencia de *GATA2* se manifiesta a través de un amplio espectro de mutaciones, con más de 150 variantes distintas identificadas hasta la fecha. Mutaciones recurrentes incluyen los cambios *missense* T354M, R398W y R396W. Las mutaciones abarcan variantes nulas o *missense*, cambios en regiones no codificantes y reguladoras, e incluso mutaciones sinónimas que afectan el procesamiento de ARN y contribuyen a la deficiencia de *GATA2*. A pesar de las numerosas mutaciones identificadas, los estudios funcionales siguen siendo limitados. En este proyecto, nuestro enfoque es emplear ensayos *in vitro* para evaluar el impacto de las mutaciones del exón 5 del gen *GATA2* en el *splicing* del ARN, contribuyendo a una comprensión más profunda de la patogénesis de la deficiencia de *GATA2*.

PALABRAS CLAVE

Neoplasias mieloides, *GATA2*, G2DS, *splicing*, mutagénesis dirigida de sitio, estudios funcionales

Alumno: Pablo López Nieto

Tutora académica: Dra. Esther Giraldo Reboloso

Tutor experimental: Dr. Alessandro Liquori

Cotutora experimental: Dra. Cristina Martínez Valiente

Valencia, julio de 2024

ABSTRACT

Myeloid neoplasms (MN) with a germline predisposition are increasingly recognized, constituting an estimated 10% of cases of myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML). Germline mutations in genes such as *GATA2*, *ETV6*, *DDX41*, *SAMD9*, and *SAMD9L* are implicated in this predisposition. *GATA2* deficiency, a rare genetic disorder resulting from mutations in the *GATA2* gene, significantly impacts the development and function of blood and immune system cells. Characterized by immunodeficiency, bone marrow failure, MDS, lung disease, lymphedema, and heightened susceptibility to infections and cancers, *GATA2* deficiency manifests through a broad spectrum of mutations, with over 150 distinct variations identified to date. Notable recurrent mutations include T354M, R398W, and R396W. The mutations encompass null mutations, missense mutations, noncoding substitutions in regulatory regions, and even synonymous mutations that disrupt RNA processing and contribute to *GATA2* deficiency. Despite the numerous mutations identified, functional studies remain limited. In this project, our focus is to employ *in vitro* assays to assess the impact of mutations located in exon 5 of the *GATA2* gene on splicing, contributing to a deeper understanding of *GATA2* deficiency pathogenesis.

KEYWORDS

Myeloid neoplasms, *GATA2*, G2DS, splicing, site-directed mutagenesis, functional studies

AGRADECIMIENTOS

En primer lugar, me gustaría agradecerle a mi tutora académica el papel que ha tenido en mi formación profesional. Esther, muchísimas gracias por acercarme la ciencia desde el pensamiento crítico, tu propio amor por lo que haces y esa cercanía que te caracteriza. Estoy enormemente feliz por encontrarme en el camino docentes como tú. De corazón, gracias.

A mis tutores experimentales, Alessandro y Krys, me habéis hecho sentir como en casa desde el primer día que entré por la puerta del laboratorio 4.08 para hacer las prácticas de la carrera. Alessandro, muchísimas gracias por tu cercanía, tu sentido del humor y por siempre buscar ampliar los horizontes de mi formación. Cualquiera día me haré millonario porque habré puesto a la venta todas las anotaciones que me has escrito en hojas de papel explicándome el proyecto (jajajaja). Krys, creo que no tengo suficientes palabras para agradecerte tu paciencia conmigo. Siempre has tenido unas palabras amables cuando las cosas no han salido como queríamos y no exagero si te digo que la experiencia que he adquirido estos dos años es, en gran parte, culpa tuya. Siempre dices que te encanta la docencia y, se nota que lo haces con gusto. No puedo olvidarme en este párrafo de mis compañeras de laboratorio María, Bea y Lucía. Mil gracias a las tres por vuestras anécdotas, vuestra ayuda y por los ratos tan buenos que hemos pasado hablando de cualquier cosa. Tengo que agradecerte especialmente a ti, Lucía, que me has guiado durante todo el TFG con tu experiencia previa, tus protocolos de valor incalculable y que me has sacado de algún apuro en alguna ocasión. Eso sí, creo que el lunes toca otro Western (jajajaja, ¡muchas gracias!).

El último pedacito de este trozo de papel digital quiero dedicárselo a mis amigos y familia. Lo primero, un millón de gracias a mis padres por permitirme el lujo de estudiar sin preocupaciones, por allanarme el camino desde que tengo uso de razón. Sois mi ejemplo de cómo el esfuerzo incansable por alcanzar un objetivo tiene su recompensa. Gracias, mamá. Gracias, papá.

A mis compañeros de carrera y de vida, Marta M, Marta F, Júlia, Miguel, Lluna, Patri, Flor, gracias por todos los momentos que hemos compartido. Habéis definido mi historia. Desde las tardes de biblio la víspera de un examen, hasta las fiestas, las comidas, los viajes. Sin todos estos recuerdos, esta etapa no tendría ningún sentido. En especial, quiero darle las gracias a mi Marta M, por las visitas a tu querido pueblo, por las cenas en tu piso (que alguna que otra fiesta nos han apañado, jajaja) y un largo etcétera. Al final, ese “entramos juntos y saldremos juntos”, se ha hecho realidad. Me dejo la Marta más importante para el final (lo siento por el resto de las Martas). Cuando la gente escucha que tengo una hermana melliza y que estudiamos la misma carrera, se sorprenden. Yo solo puedo decir que nada habría sido lo mismo sin esas noches antes de los exámenes de risa floja por el cansancio de estudiar, sin los interminables trayectos de metro y sin la combinación “melli” en las prácticas de laboratorio. Gracias por aguantarme, soy un pesado :)

Hoy, tras los 4 años más intensos de toda mi vida, entiendo mejor que nunca el significado de vivir el camino bien acompañado.

En una palabra,

Gracias.

ÍNDICE MEMORIA DESCRIPTIVA

1.- INTRODUCCIÓN	1
1.1.- LAS NEOPLASIAS HEMATOLÓGICAS DE TIPO MIELOIDES	1
1.2.- EL FACTOR DE TRANSCRIPCIÓN GATA2	3
1.3.- ALTERACIONES GENÉTICAS Y EL SISTEMA DE CLASIFICACIÓN ACMG/AMP	5
1.4.- EL SÍNDROME DE DEFICIENCIA EN GATA2	6
2.- HIPÓTESIS Y OBJETIVOS	7
3.- MATERIALES Y MÉTODOS	8
3.1.- ANÁLISIS <i>IN SILICO</i> DE VARIANTES GENÉTICAS EN GATA2	8
3.2.- SOBREENPRESIÓN DE GATA2 EN CÉLULAS HEK293	8
3.2.1.- <i>Diseño de un vector con el gen GATA2 silvestre</i>	8
3.2.2.- <i>Validación del plásmido</i>	9
3.2.2.1- <i>Diseño de cebadores sobre el gen quimérico</i>	10
3.2.2.2.- <i>Secuenciación de la región plasmídica con GATA2</i>	11
3.2.2.3.- <i>Transfección celular de HEK293 con el vector</i>	11
3.2.2.4.- <i>Extracción de ARN y proteínas de los cultivos transfectados</i>	12
3.2.2.5.- <i>Retrotranscripción y análisis de splicing por PCR</i>	13
3.2.2.6.- <i>Limpieza del producto de PCR y secuenciación de los amplicones</i>	14
3.2.2.7.- <i>Western blot a partir del extracto de proteínas del lisado celular</i>	14
3.3.- MUTAGÉNESIS DIRIGIDA PARA EL ESTUDIO DE VARIANTES GENÉTICAS	15
3.3.1.- <i>Diseño de cebadores para mutagénesis puntual</i>	15
3.3.2.- <i>PCR de mutagénesis y digestión del plásmido parental</i>	15
3.3.3.- <i>Transformación bacteriana y selección en placa</i>	16
3.3.4.- <i>Amplificación, aislamiento y purificación de los plásmidos mutantes</i>	17
3.3.5.- <i>Secuenciación de la región con la variante introducida</i>	17
3.3.6.- <i>Sobreenpresión del gen quimérico mutado en células HEK293</i>	17
3.4.- CUANTIFICACIÓN DE IMÁGENES Y ANÁLISIS ESTADÍSTICOS	18
4.- RESULTADOS	18
4.1.- RESULTADO DEL ANÁLISIS <i>IN SILICO</i> CON EL PREDICTOR SPLICEAI.....	18
4.2.- EFECTO DE LA TRANSFECCIÓN DE CÉLULAS HEK293 CON EL PLÁSMIDO SILVESTRE	19
4.3.- EFECTO DE LA SOBREENPRESIÓN DE GATA2 MUTADO EN CÉLULAS HEK293.....	23
5.- DISCUSIÓN	30
5.1.- USO DEL VECTOR PRP[Exp]-CMV>{GATA2+EGFP} PARA EL ESTUDIO DE VARIANTES GENÉTICAS EN LAS REGIONES INTRÓNICAS DE GATA2	30
5.2.- INTRODUCCIÓN DE VARIANTES GENÉTICAS MEDIANTE PCR DE MUTAGÉNESIS DIRIGIDA DE SITIO CON CEBADORES SOLAPANTES	31
5.3.- ESTUDIO DEL EFECTO DE LAS VARIANTES GENÉTICAS SOBRE LA EXPRESIÓN DEL GEN GATA2 EN CÉLULAS HEK293.....	33
6.- CONCLUSIONES	35

ÍNDICE DE FIGURAS

FIGURA 1. REPRESENTACIÓN GRÁFICA DEL GEN <i>GATA2</i>	3
FIGURA 2. MAPA DEL VECTOR VB231129-1193HBQ	9
FIGURA 3. DISEÑO DE CEBADORES PARA SECUENCIACIÓN POR SANGER DEL GEN QUIMÉRICO <i>GATA2</i>	10
FIGURA 4. PROGRAMA DE TERMOCICLADOR EMPLEADO PARA LA AMPLIFICACIÓN DEL ADNC .	13
FIGURA 5. PROGRAMA DE TERMOCICLADOR BASE PARA LA PCR DE MUTAGÉNESIS.....	16
FIGURA 6. IMÁGENES DE MICROSCOPIA DE CONTRASTE DE FASES Y DE FLUORESCENCIA (PAULA) DE CÉLULAS HEK293	20
FIGURA 7. RESULTADO DE LA AMPLIFICACIÓN POR PCR DEL ADNC MEDIANTE: (A) CEBADORES ESPECÍFICOS DEL CONSTRUCTO Y (B) CEBADORES ESPECÍFICOS DE <i>GATA2</i>	21
FIGURA 8. RESULTADO DE LA SECUENCIACIÓN POR SANGER DEL ADNC DE <i>GATA2</i> (CEBADOR EGFP_F).....	21
FIGURA 9. RESULTADO DE LA SECUENCIACIÓN POR SANGER DEL ADNC DE <i>GATA2</i> (CEBADOR E46_F)	22
FIGURA 10. RESULTADO DEL WESTERN BLOT REALIZADO A PARTIR DE LOS EXTRACTOS DE PROTEÍNA DE LOS CULTIVOS CELULARES	22
FIGURA 11. ELECTROFEROGRAMAS DE SECUENCIACIÓN MEDIANTE SANGER.....	23
FIGURA 12. IMÁGENES DE MICROSCOPIA DE CONTRASTE DE FASES Y DE FLUORESCENCIA DE CÉLULAS HEK293	24
FIGURA 13. COMPARATIVA DE LOS VALORES DE ÁREA DE CAMPO VERDE Y DE CAMPO CLARO (RATIOS F/CC) DE LOS CULTIVOS CELULARES TRANSFECTADOS	25
FIGURA 14. COMPARATIVA DE LOS VALORES DE ÁREA DE CAMPO VERDE Y DE CAMPO CLARO (RATIOS F/CC) DE LAS VARIANTES ESTUDIADAS AGRUPADAS	26
FIGURA 15. ESQUEMA DEL PATRÓN DE <i>SPLICING</i> SILVESTRE DE <i>GATA2</i>	27
FIGURA 16. AMPLIFICACIÓN POR PCR CON LOS CEBADORES E46_F/E46_R DE ADNC.....	28
FIGURA 17. WESTERN BLOT CON LOS EXTRACTOS DE PROTEÍNA DE LOS CULTIVOS CELULARES .	29

ÍNDICE DE TABLAS

TABLA 1. REACCIONES DE SECUENCIACIÓN PARA LA COMPROBACIÓN DEL ORF QUIMÉRICO	11
TABLA 2. COMPONENTES DE LOS GELES <i>RESOLVING</i> Y <i>STACKING</i> DEL SDS-PAGE.....	14
TABLA 3. PREDICCIÓN DEL IMPACTO DE LAS VARIANTES EN LOS SITIOS DE <i>SPLICING</i> DEL ALGORITMO SPLICEAI.....	19
TABLA 4. RATIO F/CC CALCULADO PARA CADA CULTIVO TRANSFECTADO	25

ACRÓNIMOS Y ABREVIATURAS

NM: neoplasia mieloide

HSC: célula madre hematopoyética (del inglés, *haematopoietic stem cell*)

MDS: síndrome mielodisplásico

LMA: leucemia mieloide aguda

GATA2: factor de unión a guanina-adenina-timina-adenina 2

CEBPA: proteína de unión a *enhancer* CCAAT alfa (del inglés, *CCAAT/enhancer-binding protein alpha*)

RUNX1: factor de transcripción relacionado con Runt 1 (del inglés, *Runt-related transcription factor*)

DDX41: helicasa 41 de caja aspártico-glutámico-alanina-aspártico (del inglés, *DEAD-box helicase 41*)

ANKRD26: proteína 26 que contiene el dominio repetido de anquirina (del inglés, *ankyrin repeat domain containing 26*)

TF: factor de transcripción (del inglés, *transcription factor*)

ZF: dedo de zinc (del inglés, *zinc finger*)

TSS: sitio de inicio de la transcripción (del inglés, *transcription start site*)

WGS: secuenciación completa del genoma (del inglés, *whole genome sequencing*)

ACMG/AMP: Colegio Americano de Genética Médica y Genómica y la Asociación para la Patología Molecular (del inglés, *American College of Medical Genetics and Genomics/Association for Molecular Pathology*)

G2DS: síndrome de deficiencia de GATA2 (del inglés, *GATA2 deficiency syndrome*)

ADN: ácido desoxirribonucleico

VUS: variante de significado incierto (del inglés, *variant of unknown significance*)

ARNm: ácido ribonucleico mensajero

ORF: marco abierto de lectura (del inglés, *open reading frame*)

ADNc: ADN complementario

EGFP: proteína verde fluorescente mejorada (del inglés, *enhanced green fluorescent protein*)

HEK293: células embrionarias humanas de riñón 293 (del inglés, *human embryonic kidney*)

SFB: suero fetal bovino

DMEM: Medio de Eagle modificado de Dulbecco (del inglés, *Dulbecco's Modified Eagle Medium*)

PBS: tampón fosfato salino (del inglés, *phosphate buffer saline*)

RIPA: tampón para ensayo de radioinmunoprecipitación (del inglés, *radioimmunoprecipitation assay buffer*)

EDTA: ácido etilendiaminotetraacético

ARNasa: hidrolasa de ácido ribonucleico

PCR: reacción en cadena de la polimerasa (del inglés, *polymerase chain reaction*)

TBE: tampón Tris, Borato y EDTA

dNTP: deoxinucleótido trifosfato

BSA: albúmina de suero bovino (del inglés, *bovine serum albumin*)

SDS: dodecilsulfato de sodio (del inglés, *sodium dodecylsulphate*)

PAGE: electroforesis en gel de poliacrilamida (del inglés, *polyacrylamide gel electrophoresis*)

dH₂O: agua ultrapura libre de nucleasas

IgG: inmunoglobulina G

HRP: peroxidasa de rábano (del inglés, *horseradish peroxidase*)

ADNp: ácido desoxirribonucleico plasmídico

SNV: variante de un único nucleótido (del inglés, *single nucleotide variant*)

NTC: control sin ADN molde (del inglés, *no template control*)

RT-PCR: retrotranscripción por reacción en cadena de la polimerasa

SDM: mutación dirigida de sitio (del inglés, *site-directed mutagenesis*)

RNP: ribonucleoproteína

Anejo III. Relación del trabajo con los Objetivos de Desarrollo Sostenible de la agenda 2030 Anexo al Trabajo de Final de Grado

A. Indicar el grado de relación del trabajo con los Objetivos de Desarrollo Sostenible (ODS).

	Alto	Medio	Bajo	No procede
ODS 1. Fin de la pobreza				x
ODS 2. Hambre cero				x
ODS 3. Salud y bienestar	x			
ODS 4. Educación de calidad				x
ODS 5. Igualdad de género				x
ODS 6. Agua limpia y saneamiento				x
ODS 7. Energía asequible y no contaminante				x
ODS 8. Trabajo decente y crecimiento económico				x
ODS 9. Industria, innovación e infraestructuras				x
ODS 10. Reducción de las desigualdades		x		
ODS 11. Ciudades y comunidades sostenibles				x
ODS 12. Producción y consumo responsables				x
ODS 13. Acción por el clima				x
ODS 14. Vida submarina				x
ODS 15. Vida de ecosistemas terrestres				x
ODS 16. Paz, justicia e instituciones sólidas				x
ODS 17. Alianzas para lograr objetivos.				x

B. Describir brevemente la alineación del TFG con los ODS, marcados en la tabla anterior, con un grado alto.

El presente TFG se enmarca bajo los ODS 3 (Salud y bienestar) con un grado alto de alineación.

En el marco del ODS 3, el objetivo central de este trabajo es el estudio de la predisposición que otorgan ciertas variantes genéticas de origen germinal al desarrollo de un síndrome raro conocido como deficiencia de *GATA2*, que aumenta la susceptibilidad ante neoplasias mieloides. Sin embargo, el conocimiento acerca de los mecanismos patogénicos es muy limitado, por lo que el aumento de la información a nivel molecular que conduce al desarrollo de una neoplasia hematológica de tipo mieloides es un factor que puede ayudar al diagnóstico y al posterior desarrollo de una terapia farmacológica dirigida. Con ello, se pretende "Garantizar una vida sana y promover el bienestar para todos en todas las edades", tal y como queda reflejado en este Objetivo de Desarrollo Sostenible.

1.- INTRODUCCIÓN

1.1.- LAS NEOPLASIAS HEMATOLÓGICAS DE TIPO MIELOIDES

El sistema hematopoyético humano comprende un conjunto muy heterogéneo de tipos celulares: los eritrocitos, que son responsables del transporte de oxígeno y dióxido de carbono; las plaquetas, que actúan en la coagulación sanguínea y en la reparación de tejidos ante traumatismos vasculares; y los leucocitos, encargados de coordinar y efectuar la inmunidad celular y humoral ante elementos reconocidos como extraños (Haas et al., 2018; Peters et al., 2023). Todos ellos comparten un progenitor común: las células madre hematopoyéticas (HSC) (Peters et al., 2023), que se alojan en la médula ósea en un estado indiferenciado (Yu et al., 2016). El proceso biológico mediante el cual las HSC dan lugar a los componentes celulares de la sangre se denomina hematopoyesis e involucra la diferenciación y proliferación de estas células multipotentes en subtipos celulares bien definidos (da Silva et al., 2023). Bajo este preámbulo, el modelo más conocido mediante el cual una célula indiferenciada alcanza un estado comprometido es la teoría epigenética de Waddington. Esta teoría describe la diferenciación celular como un *paisaje montañoso* en el que la célula progenitora *es una pelota que va rodando cuesta abajo*, pasando por *valles intermedios*, hasta que encuentra un *valle final del cual no puede salir*, que representa su estado más estable y que constituye un tipo celular bien definido (Feinberg et al., 2023; Ferrell et al., 2012; Waddington et al., 1942). Esta metáfora de la diferenciación celular, aunque representa solamente el papel de la epigenética, ilustra la compleja regulación que ejercen unos productos génicos sobre otros bajo la influencia del entorno para dar lugar a un fenotipo concreto (Tronick et al., 2016).

En esta dirección, en un sistema hematopoyético sin alteraciones, las HSC se diferencian y se expanden clonalmente en su progenie madura. Este proceso se regula mediante un programa transcripcional codificado en el genoma, que se modifica progresivamente por factores principalmente epigenéticos, hecho que define las características finales de la población celular. Sin embargo, las células progenitoras pueden volverse incapaces de diferenciarse adecuadamente o proliferar de manera descontrolada si se producen alteraciones en los genes que controlan la diferenciación, la proliferación o la muerte celular programada (Feinberg et al., 2023). De hecho, estos rasgos corresponden con algunos de los sellos distintivos definidos por Hanahan & Weinberg (2000) para definir a las células cancerosas.

En el contexto de la hematopoyesis aberrante, las neoplasias mieloides (NM) comprenden un conjunto de enfermedades genéticas que afectan a 1 de cada 10.000 individuos cada año y que se caracterizan por el colapso funcional de las células del linaje mieloides, como consecuencia de la pérdida del control proliferativo y de la diferenciación de sus progenitores hematológicos (Churpek et al., 2019; Gu et al., 2023; Sahoo et al., 2020). La Organización Mundial de la Salud (OMS) define 5 grandes grupos para clasificar a estas patologías en su revisión de 2022 (Khoury et al., 2022): (1) las neoplasias mieloproliferativas (NMP), (2) las neoplasias mielodisplásicas (MDS), (3) las neoplasias mieloproliferativas/mielodisplásicas (NMP/SMD), (4) las leucemias mieloides agudas (LMA) y, por último, (5) las NM secundarias, que aparecen como resultado de la exposición a

agentes citotóxicos tras un tratamiento anterior y requieren ser consideradas como un grupo independiente. Generalmente, estas enfermedades aparecen de manera esporádica en pacientes de edad avanzada (de >65 años de media) y el riesgo aumenta con la edad de cada individuo (Rafei et al., 2019). Sin embargo, la identificación de alteraciones germinales en los genes *RUNX1* y *CEBPA* en pacientes afectados por NM abrió la puerta al descubrimiento de síndromes de carácter monogénico, en los que se relacionaron alteraciones genéticas germinales en heterocigosis en genes involucrados en la génesis y diferenciación de los progenitores hematológicos con una susceptibilidad aumentada al desarrollo de NM (Churpek et al., 2019; Gu et al., 2023; Rafei et al., 2019; Sahoo et al., 2020; Santiago et al., 2023). De hecho, desde 2016, la OMS ha establecido una nueva categoría que engloba este conjunto de patologías bajo el término “neoplasias mieloides de predisposición germinal”, que involucra NM derivadas de mutaciones en genes como *GATA2*, *CEBPA*, *RUNX1*, *DDX41* o *ANKRD26*, entre otros (Arber et al., 2016; Arber et al., 2022; Khoury et al., 2022). En este contexto, se estima que entre un 10-15% de los pacientes pediátricos y jóvenes diagnosticados de MDS o LMA podrían portar alteraciones de origen germinal (Rafei et al., 2019), lo que es consistente con que hasta un 75% de los pacientes afectados por estos síndromes desarrollan NM con una edad mediana de 20 años (Homan et al., 2021; Sahoo et al., 2020).

No obstante, estos síndromes presentan grandes dificultades diagnósticas. En primer lugar, el espectro de manifestaciones clínicas asociado de forma específica a las alteraciones moleculares es limitado (Hirabayashi et al., 2018; Sahoo et al., 2020; Santiago et al., 2023). Además, incluso en familias con múltiples individuos que presentan las mismas variantes germinales, existen pacientes asintomáticos o sin un historial familiar informativo. La penetrancia incompleta y, en algunos casos, la latencia complica aún más el manejo de estas neoplasias (Rajput et al., 2023; Sahoo et al., 2020; Oleaga-Quintas et al., 2021). De hecho, en muchas ocasiones, los pacientes afectados por estos síndromes no exhiben manifestaciones clínicas previas al desarrollo de la NM, lo que impide contar con un historial clínico que ayude a guiar el diagnóstico. A pesar de ello, en algunos casos, se han podido identificar ciertos patrones moleculares y clínicos que facilitan el diagnóstico a un ojo experto. A nivel molecular, por ejemplo, encontramos que los pacientes con variantes germinales de *DDX41* suelen adquirir mutaciones somáticas en el mismo gen (fenómeno del “segundo hito”) o variantes *missense* en el gen *GATA2* que se ubican en un *hotspot* mutacional en el dominio de dedos de zinc 2 (ZF2) (Calvo et al., 2023; Rajput et al., 2023). Asimismo, existen algunos signos clínicos que específicos de síndromes: muchos pacientes con alteraciones en *RUNX1/ETV6/ANKRD26* suelen presentar un recuento bajo de plaquetas (trombocitopenia), mientras que los individuos con deficiencia en *GATA2* exhiben comorbilidades como sordera o linfedema (Kotmayer et al., 2022; Sahoo et al., 2020).

Por tanto, la caracterización de las causas genéticas de estos síndromes y su impacto de cara al desarrollo de NM es de urgencia para anticipar el diagnóstico, mejorar el manejo clínico y ampliar la estrategias terapéuticas disponibles para estas patologías.

1.2.- EL FACTOR DE TRANSCRIPCIÓN GATA2

El gen *GATA2*, situado en la banda citogenética q21.3 del cromosoma 3 (GRCh38.p14), consta de 8 exones y codifica para 2 isoformas diferentes de un factor de transcripción (TF) nuclear, con pesos moleculares de, aproximadamente, 50,5 y 49,0 kDa. Este TF pertenece a la familia conservada de factores de transcripción GATA, denominada así por la secuencia consenso de nucleótidos (A/T)GATA(A/G) que reconocen estas proteínas para unirse a regiones promotoras y/o *enhancer* en el ADN (Aktar et al., 2023; Hewitt et al., 2016; Tremblay et al., 2018). Para llevar a cabo esta unión, los miembros de esta familia de seis proteínas poseen dos dominios ZF: ZF1 (también llamado dedo N-terminal) y ZF2 (o dedo C-terminal) (Jung et al., 2018; Sahoo et al., 2020; Shimizu et al., 2020). Estos dominios les permiten promover la transcripción de genes diana involucrados en la diferenciación celular. Tradicionalmente, las proteínas GATA se han clasificado en dos grandes grupos en función de su implicación en la diferenciación de progenitores hematológicos (*GATA1*, *GATA2* y *GATA3*) o cardíaco-endoteliales (*GATA4*, *GATA5* y *GATA6*) (Aktar et al., 2023; Khandekar et al., 2007; Tremblay et al., 2018). Sin embargo, esta clasificación no tiene en cuenta el papel pleiotrópico que ejerce cada TF (Mir et al., 2015; Tremblay et al., 2018).

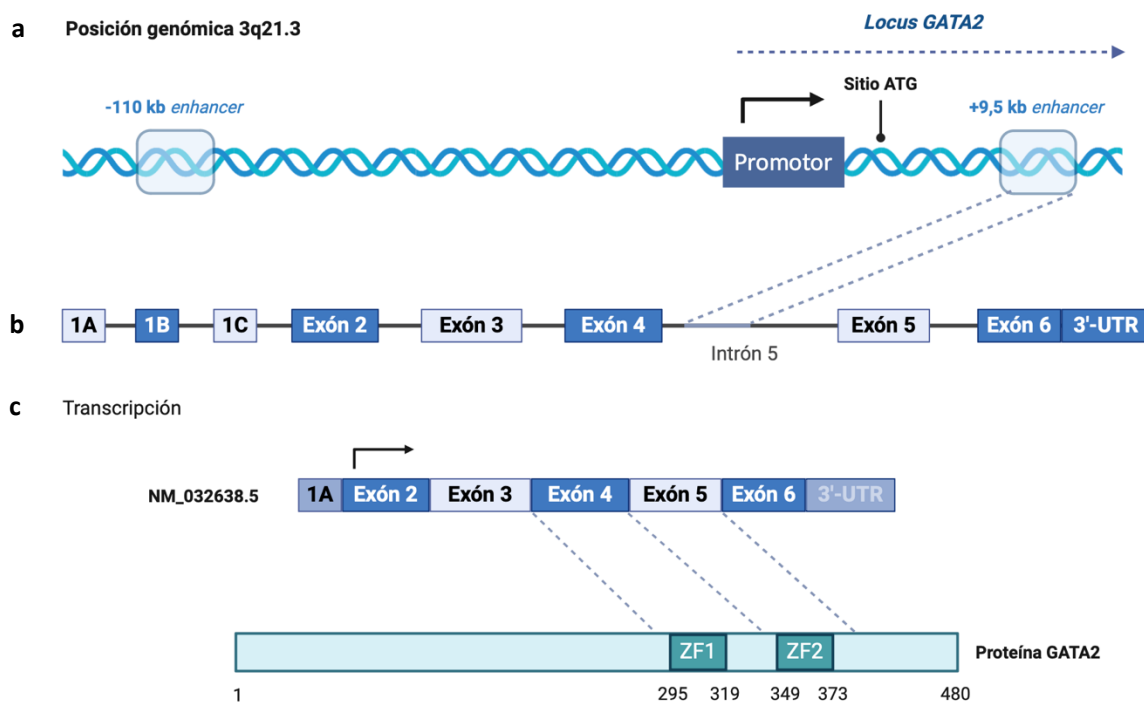


Figura 1. Representación gráfica del gen *GATA2*. a Posiciones de las regiones *enhancer* -110 kb y +9,5 kb respecto al locus de *GATA2* b Estructura del gen *GATA2* en exones c Transcrito maduro canónico y proteína *GATA2* resultante. ZF: Zinc Finger. Imagen: Created with BioRender.com.

La proteína *GATA2* ejerce su acción como un regulador central de la hematopoyesis en conjunción con los factores de transcripción *RUNX1*, *SCL/TAL1*, *LYL1*, *LMO2*, *ERG* y *FLI1*, con los que forma un complejo heptamérico que actúa sobre más de 1000 *loci* en los progenitores sanguíneos (Collin et al., 2015; Rajput et al., 2023). Desde las etapas embrionarias tempranas, *GATA2* es un componente esencial en la transición endotelio-hematopoyética (EHT), así como en la supervivencia y renovación

de las HSC que resultan de este proceso de diferenciación a partir del endotelio hemogénico (Kotmayer et al., 2023; Mir et al., 2015; Peters et al., 2023). La actividad temprana de este TF es crucial para la génesis de la mayoría de los linajes hematológicos. Así, por un lado, la alta expresión de *GATA2* en las HSC induce su autorrenovación, mientras que la disminución gradual de su expresión es necesaria para que las células hija se comprometan en un linaje (Peters et al., 2023; Shimizu et al., 2020). Por ejemplo, la diferenciación mieloeritroide, que culmina con la separación del linaje mieloide y eritroide a partir de las células madre mieloides (MSCs), tiene lugar tras la represión de la expresión de *GATA2* que ejerce *GATA1* (Churpek et al., 2019; Mir et al., 2015; Peters et al., 2023). Del mismo modo, la inducción de la expresión de *GATA2* en células comprometidas es necesaria para la maduración de ciertos tipos celulares de linaje mieloide, como los megacariocitos, así como para la diferenciación de basófilos y mastocitos (Collin et al., 2015; Rajput et al., 2023). Por lo tanto, la expresión de este TF está estrictamente ligada al contexto celular, hecho que remarca su control preciso y dinámico (Churpek et al., 2019; Mehta et al., 2017; Shimizu et al., 2020).

Esta regulación compleja de la expresión del TF se debe en gran medida a la existencia de múltiples elementos de regulación transcripcional en cis en el *locus* del gen *GATA2* (Aktar et al., 2023). De importancia crítica para la hematopoyesis son los *enhancers* -110 Kb y +9,5 Kb. Los *enhancers* se nombran por su posición relativa respecto al sitio de inicio de la transcripción (TSS) (Friedman et al., 2024), aplicándose el mismo criterio para gen *GATA2* humano. La **figura 1** recoge una representación gráfica del gen *GATA2*, uno de sus dos transcritos canónicos (NM_032638) así como del producto proteínico.

Los *enhancers* son regiones genéticas cortas (50-500 pb), situadas más de 1 kb del TSS que contienen motivos de ADN para el reclutamiento de factores de transcripción con funciones específicas de linaje celular (Friedman et al., 2024; Hewitt et al., 2016). Por un lado, el *enhancer* -110 kb, que en la literatura se suele mencionar indistintamente como el *enhancer* -77 kb murino por ser en el genoma de este el modelo animal donde se identificó dicha región por primera vez (Collin et al., 2015), es un elemento distal aguas arriba del *locus* de *GATA2* mediante el que *GATA2* autorregula su propia expresión en los progenitores mieloides comunes, los progenitores de granulocitos y macrófagos, y los progenitores de megacariocitos y eritrocitos (Aktar et al., 2023; Suzuki et al., 2019). Sin embargo, su impacto en la proliferación y supervivencia de las HSC es mínimo (Churpek et al., 2019; Hewitt et al., 2016; Kotmayer et al., 2022). Varios estudios en ratones KO (del inglés, *knockout*) nulicigotos para esta región *enhancer* (-77^{-/-}) han confirmado esta actividad por el hecho de que la génesis de HSC tiene lugar con normalidad en estos roedores, pero se produce una acumulación de progenitores mieloides incapaces de diferenciarse en la progenie mieloeritroide completa (Hewitt et al., 2016; Kotmayer et al., 2022; Mir et al., 2015; You et al., 2020). Por otro lado, el *enhancer* +9,5 kb está localizado en el intrón 5 del gen *GATA2* humano (Calvo et al., 2023) y es una región suficiente a la vez que necesaria para inducir la EHT y la renovación de las HSC (Dobrzycki et al., 2020; Jung et al., 2018). Estudios previos han demostrado que su delección en homocigosis (+9,5^{-/-}) se manifiesta en la pérdida de la totalidad de la población de HSC en modelos murinos (Churpek et al., 2019; Kotmayer et al., 2022) y produce letalidad embrionaria alrededor de las 13-14 semanas (Dobrzycki et al., 2020; Mehta et al., 2017). El alto grado de conservación de esta

región *enhancer* manifiesta su protagonismo en la regulación de la hematopoyesis por la presencia de motivos de reconocimiento para varias de proteínas. En sentido 5'-3', aparecen en orden: (1) la secuencia E-box (CATCTG), motivo canónico para el TF hematopoyético SCL/TAL1; (2) la secuencia (A/T)GATA(A/G), reconocida por las proteínas GATA1 y GATA2 de manera competitiva; y (3) el motivo de reconocimiento para la superfamilia de factores de transcripción Ets (TTCC), como FLI1, ERG o ETV2 (Soukup et al., 2020).

En definitiva, tanto la acción temprana y la ubicuidad de *GATA2* como la complejidad en cuanto a la regulación de este TF pone de manifiesto el papel vital del mismo en la hematopoyesis. Por lo tanto, es fácil prever que la existencia de alteraciones que modifiquen la síntesis, la estructura o la actividad de *GATA2* pueden tener consecuencias biológicas negativas que conduzcan al desarrollo de una patología, como se describirá en apartados posteriores.

1.3.- ALTERACIONES GENÉTICAS Y EL SISTEMA DE CLASIFICACIÓN ACMG/AMP

El significativo desarrollo y evolución de las técnicas de secuenciación masiva en las últimas décadas han provocado un cambio en el paradigma de los diagnósticos genéticos. Se ha pasado de un análisis predominantemente clínico, basado en las manifestaciones fenotípicas de los pacientes, a un enfoque centrado principalmente en la genética. Hoy en día, la tendencia del sector apunta a que la secuenciación del genoma completo (WGS) marcará el futuro cercano del diagnóstico clínico (Balloux et al., 2018). Sin embargo, aunque cada vez se investiga más la posible causa genética de las enfermedades y los análisis genéticos son más accesibles, la interpretación de los resultados de secuenciación no es trivial (Richards et al., 2015). Por un lado, la cantidad de información que emerge de la secuenciación masiva del genoma o de otras técnicas como el exoma o el transcriptoma alcanza el rango de terabytes (Katsanis et al., 2013), por lo que se requiere personal y herramientas informáticas especializadas para analizar los datos e identificar los cambios (es decir, las variantes genéticas) con respecto a la secuencia del genoma de referencia. Por otro lado, la interpretación del impacto funcional de estas variantes es muy compleja, ya que solo algunas de ellas pueden tener un impacto en la enfermedad estudiada. Es necesario distinguir estas variantes de las demás polimórficas, es decir, aquellas con un impacto neutral.

En esta dirección, en el año 2015, el Colegio Americano de Genética Médica y Genómica y la Asociación para la Patología Molecular (ACMG/AMP) publicaron una serie de directrices para la clasificación de variantes genéticas puntuales (SNV) con el fin de establecer su posible impacto funcional (Richards et al., 2015). Este marco surgió a partir de la necesidad de determinar de manera estandarizada y homogénea el impacto de las variantes genéticas y, actualmente, es el sistema más utilizado. Estas recomendaciones incluyen 28 criterios de evidencia individuales que se otorgan en función de varios aspectos, como la frecuencia poblacional de la variante, el tipo de alteración y la región genómica donde se ubica o la existencia de evidencias previas en pacientes con la misma enfermedad (Harrison et al., 2019). A cada uno de estos criterios se les asigna una evidencia en favor de la benignidad de la variante (B) o de su patogenicidad (P), estableciendo 5 niveles de evidencia, que incluyen: *stand-alone* (A), *very strong* (VS), *strong* (S), *moderate* (M), o *supporting*

(P). Tras la combinación de los diferentes criterios con sus niveles de evidencia específicos, cada variante puede clasificarse en 5 categorías, como patogénica, potencialmente patogénica, de significado clínico incierto (VUS), potencialmente benigna o benigna (Harrison et al., 2019; Richards et al., 2015).

Sin embargo, en muchas ocasiones, esta clasificación está limitada por la falta información (McClure et al. 2018), que obstaculiza la aplicación de algunos de los criterios. Como consecuencia, la mayoría de las variante identificadas en un análisis de secuenciación masiva se clasifican como VUS. Por ejemplo, solo en el gen *GATA2* se han descrito más de 1500 variantes germinales, de las cuales prácticamente el 50% se enmarcan en la categoría de VUS. Es, por tanto, urgente un esfuerzo global que aporte información a nivel funcional para la reclasificación de VUS.

1.4.- EL SÍNDROME DE DEFICIENCIA EN *GATA2*

El síndrome de deficiencia de *GATA2* (G2DS) es un trastorno genético raro causado por mutaciones heterocigotas en el gen *GATA2* que predisponen al desarrollo de NM (Fabozzi et al., 2021). Hasta la fecha, entre los más de 500 pacientes reportados en la literatura con G2DS, se han descrito más de 200 alteraciones genéticas patogénicas o potencialmente patogénicas (Homan et al., 2021; Shimizu et al., 2020). Se pueden agrupar en 4 categorías: 1) Variantes truncantes en la región del ZF2. Se trata de variantes que afectan a los sitios canónicos de *splicing*, mutaciones sinsentido (*nonsense*) o que desplazan el marco de lectura (*frameshift*) y deleciones génicas; 2) Mutaciones de sentido erróneo (*missense*) en el dominio ZF2. 3) Cambios sinónimos que afectan al procesamiento del ARNm; 4) Variantes ubicadas en las regiones implicadas en la regulación de la expresión de *GATA2* (especialmente, en la región *enhancer* +9,5 kb) (Churpek et al., 2019; Santiago et al., 2023). Todo ello, sumado a la existencia de más de 700 VUS, cuya interpretación de patogenicidad es complicada, pone de manifiesto el carácter fundamentalmente genético del síndrome.

La enfermedad se origina con la inactivación monoalélica de *GATA2* que, probablemente, conduce a la haploinsuficiencia del factor de transcripción y a una hematopoyesis defectuosa. Esto se explica por la función de *GATA2* en la autorrenovación de las HSC y en la diferenciación de la progenie mieloide. Por un lado, la reducción a la mitad de la expresión de *GATA2* no es suficiente para que las HSC proliferen, sino que induce su diferenciación, lo que provoca que su depleción (Calvo et al., 2023; Collin et al. 2015). Por otro lado, en los progenitores mieloides la haploinsuficiencia del factor de transcripción impide su diferenciación definitiva (Collin et al., 2015).

Sin embargo, este mecanismo no explica completamente la heterogeneidad observada entre pacientes diagnosticados con G2DS. Previamente al desarrollo de SMD/LMA, la enfermedad se manifiesta con un cuadro sindrómico complejo (Hirabayashi et al. 2017), incluyendo inmunodeficiencias por déficit de las poblaciones de células dendríticas, monocitos, células NK linfoides (deficiencia DCML), lo que resulta en infecciones recurrentes por microorganismos oportunistas, especialmente, por complejos de *Mycobacterium avium* (conocido como síndrome MonoMAC) (Fujiwara, 2017; Homan et al., 2021; Kotmayer et al., 2022). Además, los individuos afectados pueden presentar linfedemas con mielodisplasia (síndrome de Emberger) y, en menor

medida, neutropenia crónica, proteinosis alveolar pulmonar o sordera neurosensorial, entre otras manifestaciones (Dickinson et al., 2014; Hsu et al., 2015; Shamriz et al., 2022).

Aunque el G2DS representa solo el 5-7% de los todos los casos de MDS/AML (Churpek et al., 2019; Rajput et al., 2023), hasta el 80% de los pacientes con mutaciones en *GATA2* desarrollan un cáncer hematológico antes de los 20 años (Rajput et al., 2023; Rafei et al., 2019). El origen de estas variantes difiere entre estudios: mientras que algunos afirman que lo más común es que los pacientes presenten mutaciones *de novo* (hasta un 80%) (Sahoo et al., 2020), otros estudios han encontrado que prácticamente el 50% de los casos sigue un patrón de herencia autosómico dominante y que tan solo el 5% de los pacientes afectados no tienen un historial familiar (Calvo et al., 2023; Santiago et al., 2023). Entre las mutaciones más frecuentes en *GATA2*, se encuentran la *missense* c.1061C>T (p.Thr354Met), que afecta al dominio ZF2 de *GATA2*, la variante sinónima c.351C>G (p.Thr117=) que produce alteraciones a nivel postranscripcional, la variante *nonsense* c.988C>T (p.Arg330*) que produce la terminación prematura de la proteína en la región entre los dominios ZF1 y ZF2 y el cambio c.1017+572C>T, que afecta a la región *enhancer* +9,5 kb presente en intrón 4 de *GATA2* (Homan et al., 2021; Kozyra et al., 2020)

2.- HIPÓTESIS Y OBJETIVOS

A pesar de ser uno de los genes de predisposición más estudiados, la mayoría de las variantes identificadas en *GATA2* se clasifican como VUS. Aunque muchas de ellas son privadas, es decir, se encuentran de forma aislada en pacientes o familias con G2DS, se han identificado mutaciones recurrentes o en regiones *hotspot*. A menudo, estas variantes se clasifican como patogénicas debido a su frecuencia en las cohortes de pacientes con G2DS, sin considerar los mecanismos moleculares subyacentes a su patogenicidad. Las mutaciones en *GATA2* son de diferentes tipos y, por tanto, no todas afectan la función proteica de la misma manera. Por ejemplo, variantes en la región *enhancer* +9,5 kb podrían alterar los niveles de expresión de *GATA2* sin afectar al patrón de *splicing*, mientras que cambios en los sitios canónicos de *splicing* podrían resultar en eventos de *splicing* complejo.

Nuestra hipótesis es que es posible descifrar el impacto funcional de una variante de *GATA2* considerando los diferentes mecanismos moleculares que podrían estar alterados como resultado de una mutación, incluyendo el *splicing* y la expresión proteica. Además, creemos que el uso de herramientas de análisis bioinformático, como los predictores de *splicing*, puede ayudar a priorizar las variantes que deben ser investigadas en estudios funcionales.

Así, los objetivos de este trabajo de fin de grado se resumen en cinco aspectos clave:

- 1.- Validación de un plásmido para el estudio del impacto de ciertas variantes de *GATA2* a nivel de ARN y a nivel proteico.
- 2.- Análisis *in silico* de variantes previamente descritas en la literatura.
- 3.- Realización de la mutagénesis dirigida para las variantes de interés, seleccionadas en el punto 2.

4.- Análisis del patrón de *splicing* que inducen las variantes obtenidas.

5.- Estudio del efecto de las variantes en la proteína GATA2, a nivel estructural y de su expresión celular.

3.- MATERIALES Y MÉTODOS

3.1.- ANÁLISIS *IN SILICO* DE VARIANTES GENÉTICAS EN *GATA2*

Se obtuvo una lista de 149 variantes patogénicas, potencialmente patogénicas y VUS según la clasificación ACMG/AMP a partir de un trabajo de revisión publicado por Homan et al., 2021 (**Tabla Suplementaria 1**). Este conjunto de variantes se analizó con la herramienta VEP (*Variant Effect Predictor*: <http://www.ensembl.org/Help/View?id=484>) de ENSEMBL, incorporando las predicciones sobre el *splicing* que ofrece el algoritmo SpliceAI. Se filtraron las variantes de interés utilizando un valor de corte igual o superior a 0,20 de SpliceAI. Asimismo, se cotejó la lista de variantes seleccionadas alrededor del exón 5 con el conjunto de variantes disponibles en la base de datos ClinVar. Además, para ampliar la región génica estudiada, se seleccionaron variantes ya descritas en la literatura y que afectan a la región *enhancer* +9,5, la cual se encuentra frecuentemente alterada en pacientes con G2DS. Se seleccionaron las variantes c.1017+532T>A (que produce una transversión en la secuencia E-box), c.1017+572C>T (que produce una transición en el motivo de unión para los factores de transcripción Ets) y las variantes c.1017+582G>T y c.1017+693dup. Además, se decidió incluir la variante *missense* c.1061C>T (p.Thr345Met) como control positivo de patogenicidad, siendo esta la variante germinal más frecuentemente identificada en pacientes diagnosticados de G2DS.

3.2.- SOBREEXPRESIÓN DE *GATA2* EN CÉLULAS HEK293

3.2.1.- Diseño de un vector con el gen *GATA2* silvestre

Con el objetivo de estudiar las variantes de *GATA2 in vitro*, se diseñó un plásmido que incluyera la secuencia genómica completa del gen, tomando en cuenta las limitaciones que supondrían emplear constructos genéticos de tamaño superior. Para ello, se empleó la herramienta de diseño de vectores de VectorBuilder (<https://en.vectorbuilder.com/design.html>).

En primer lugar, se diseñó el marco abierto de lectura (ORF, del inglés, *open reading frame*) valorando la posibilidad de emplear la secuencia génica completa de *GATA2*. Para ello, se descargó la secuencia RefSeq de la isoforma canónica de *GATA2* (NM_032638.5) desde su entrada en el buscador genómico ENSEMBL (ENST00000341105.7). Sin embargo, el tamaño completo de este gen (alrededor de 14 Kb) complicaría considerablemente la manejabilidad del constructo con técnicas de biología molecular estándares. Por ello, se diseñó un “mini” gen que incluía 250 pb de cada extremo de los intrones 1-2, 2-3 y 3-4 y la secuencia intrónica completa para los intrones más pequeños (4-5 y 5-6). De esta manera, se obtuvo un ORF de aproximadamente 8 Kb (**Anexo 1**). Además, se incluyó la secuencia de ADNc de la proteína verde fluorescente mejorada (EGFP) en el

final de la región 5'-UTR de *GATA2*, específicamente, en el exón 2, como forma de monitorizar la localización celular y las cantidades del producto proteínico.

A continuación, se seleccionó el promotor inmediato del citomegalovirus humano (CMV) por ser un promotor fuerte en la mayoría de los tipos celulares. Finalmente, se mantuvieron por defecto el gen de resistencia a ampicilina, la secuencia de iniciación de la traducción Kozak, la secuencia de poliadenilación tardía del virus simio 40 (SV40, del inglés, *simian virus 40*), así como el origen de replicación pUC de elevado número de copia (**Figura 2**).

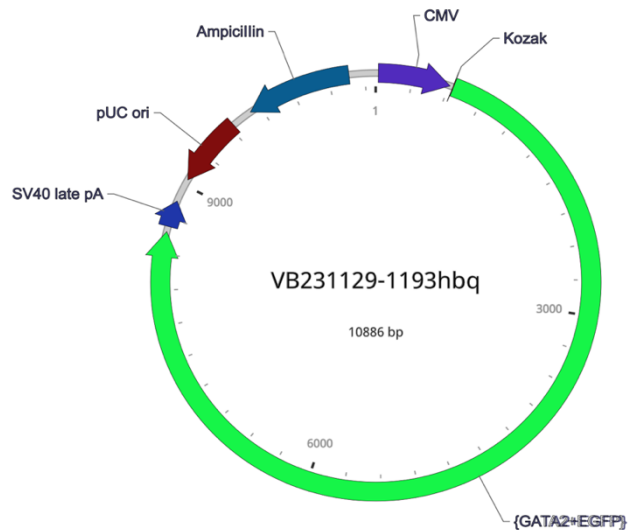


Figura 2. Mapa del vector VB231129-1193hbq. CMV: Promotor/Enhancer temprano del citomegalovirus humano; {GATA2+EGFP}: ORF del vector Kozak: Secuencia de iniciación de traducción; SV40 late pA: señal de poliadenilación tardía del virus simio 40; pUC ori: origen de replicación pUC; Ampicillin: gen de resistencia al antibiótico ampicilina.

Se solicitó el servicio de clonaje de VectorBuilder para la síntesis de este vector (pRP[Exp]-CMV>{GATA2+EGFP}). La empresa entregó el plásmido en la cepa VB UltraStable de *E. coli*. Adicionalmente, se encargó una preparación de plásmido purificado (MaxiPrep) a esta misma casa comercial. Este servicio nos permitió disponer de una suspensión de plásmido en una concentración de 1.825 ng/ μ L, a partir de la cual se extrajeron alícuotas de 50 ng/ μ L.

3.2.2.- Validación del plásmido

Para confirmar que el vector solicitado se había clonado correctamente, se secuenció la MaxiPrep y se transfectó en un cultivo celular de HEK293 (del inglés, *human embryonary kidney*). Este se observó mediante microscopía de fluorescencia PAULA (Leica) para detectar la proteína de fusión EGFP-GATA2. Posteriormente, se realizó una extracción de ARN (*RNeasy Mini Kit*, QIAGEN) y proteínas del cultivo celular (método de precipitación con acetona). A partir del ARN, se obtuvo ADNc por retrotranscripción (*PrimeScript RT Reagent Kit*, Takara) y se realizaron amplificaciones por PCR con cebadores específicos del vector y con cebadores específicos de *GATA2*. Con la extracción proteica, se buscó ratificar la masa molecular de la proteína de fusión EGFP-GATA2 mediante un *Western blot*.

3.2.2.1- Diseño de cebadores sobre el gen quimérico

En primer lugar, se diseñaron cebadores para la secuenciación Sanger del gen quimérico (EGFP+GATA2), teniendo en cuenta que se trataba de una secuencia de ADN. Se diseñaron cebadores de una longitud aproximada de 20 pb, con un porcentaje de G+C cercano al 60% y evitando cebadores que fuesen autocomplementarios, que formasen dímeros con otros cebadores o estructuras secundarias. Además, se buscaron secuencias con una temperatura de fusión cercana a 60 °C. Para evaluar estos parámetros, se emplearon las herramientas en línea *OligoAnalyzer™ Tool* (Integrated DNA Technologies, IDT) y *Multiple Primer Analyzer* (Thermo Fisher Scientific). Puesto que la secuenciación Sanger genera lecturas de aproximadamente 1.000 pb, se consideró este tamaño para cubrir toda la extensión de la región a secuenciar. La **Tabla Suplementaria 2** recopila dicho conjunto de datos para todos los cebadores de manera explícita.

Específicamente para la validación de variantes, se diseñaron dos pares de cebadores en la región que abarca desde el intrón 4-5 hasta el intrón 5-6. En total, se diseñaron 17 cebadores diferentes para cubrir la región del vector con el gen (**Figura 3**).

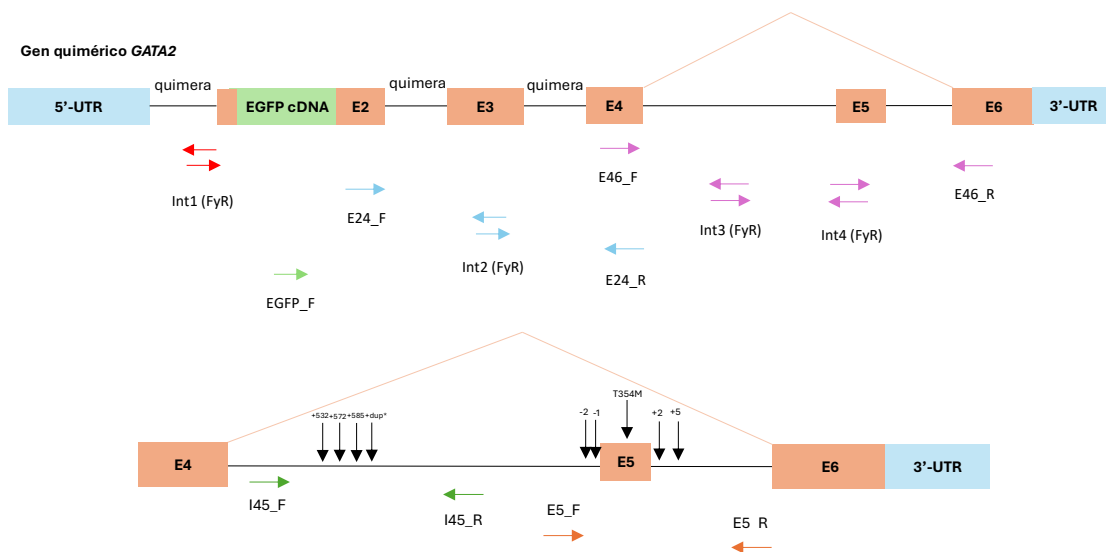


Figura 3. Diseño de cebadores para secuenciación por Sanger del gen quimérico GATA2. En rectángulos de color naranja se representan los exones de GATA2; en rectángulos de color azul las regiones codificantes no traducidas (UTR); las líneas negras representan los intrones; las flechas horizontales indican la posición y dirección de los cebadores; las flechas verticales de color negro indican la posición de las variantes de interés seleccionadas en este estudio.

Se solicitó la síntesis de estos oligonucleótidos a la empresa IDT. Los oligonucleótidos liofilizados recibidos se resuspendieron en agua ultrapura Milli-Q libre de nucleasas (dH₂O), generando disoluciones stock a 100 µM. De estas disoluciones maestras se realizaron diluciones 1:10 para producir alícuotas de uso frecuente (10 µM). Tanto el stock como las alícuotas de uso común se guardaron en frío (-20 °C) para su correcta conservación.

3.2.2.2.- Secuenciación de la región plasmídica con GATA2

Se prepararon las reacciones de Sanger individuales, tomando 5 µL de una dilución de plásmido a 50 ng/µL (250 ng de DNA plasmídico), 5 µL de dH₂O y 3 µL de la dilución a 10 µM de cada cebador. La secuenciación del gen se encargó al servicio de secuenciación Sanger de la empresa STABVIDA. En total, se enviaron 17 reacciones, cada una con un cebador diferente (**Tabla 1**).

Tabla 1. Reacciones de secuenciación para la comprobación del ORF quimérico

Reacción	Cebador	Secuencia de nucleótidos	Región de hibridación
1	Int1_F	5'-CATGGTGAGCAAGGGCGAG-3'	Exón 2
2	Int1_R	5'-CTCGCCCTTGCTCACCATG-3'	Exón 2
3	Int2_F	5'-CGTCTCCAGCCTCATCTTC-3'	Exón 3
4	Int2_R	5'-GAAGATGAGGCTGGAGACG-3'	Exón 3
5	Int3_F	5'-GAGTCAGAATTCCAGCTTCAG-3'	Intrón 4-5
6	Int3_R	5'-CTGAAGCTGGAATTCTGACTC-3'	Intrón 4-5
7	Int4_F	5'-CTCCTTGACTGAGCTGGTG-3'	Intrón 4-5
8	Int4_R	5'-CACCACTCAGTCAAGGAG-3'	Intrón 4-5
9	EGFP_F	5'-ACAACACTACAACAGCCACAACG-3'	ADNc de EGPF
10	E24_F	5'-CTGGCGCACAACACTACATGG-3'	Exón 2
11	E24_R	5'-GATGAGTGGTCGGTTCTGC-3'	Exón 4
12	I45_F	5'-GGAGCTGGCGGTTAATTACAG-3'	Intrón 4-5
13	I45_R	5'-TCTGAGTCCTAGTCCCTCAG-3'	Intrón 4-5
14	E46_F	5'-CGGGAGTGTGTCAACTGTG-3'	Exón 4
15	E46_R	5'-CACCATGCTGGACGGGTG-3'	Exón 6
16	E5_F	5'-CAGCTGCTACAGTGTGGAC-3'	Intrón 4-5
17	E5_R	5'-GTCAGTGGCCTGTTAACCTA-3'	Intrón 5-6 y Exón 6

Los resultados de secuenciación se visualizaron con el programa *4Peaks*, que es capaz de abrir archivos con extensión “.abi”. Las secuencias recibidas se comprobaron manualmente con la secuencia quimérica diseñada de *GATA2*. Además, se realizó un alineamiento de las lecturas de Sanger frente a la secuencia completa del gen quimérico (**Figura suplementaria 1**), empleándose la versión 7 en línea del programa MAFFT (del inglés, *Multiple Alignment using Fast Fourier Transform*, <https://mafft.cbrc.jp/>). Se mantuvieron los parámetros por defecto y se ejecutó el alineamiento, obteniéndose un archivo .fasta con las lecturas Sanger alineadas al constructo original. Este archivo se visualizó gráficamente empleando la herramienta NCBI MSA Viewer (versión 1.25.0). Para comprobar la secuencia a nivel nucleotídico, se empleó el software MEGA11.

3.2.2.3.- Transfección celular de HEK293 con el vector

Para la transfección, se sembraron células de la línea HEK293 en placas de 6 pocillos. Para ello, primeramente, se obtuvieron las células a partir de un cultivo previo subconfluente de pase 5 en placa T75. Este se levantó mediante digestión con tripsina, se recuperaron las células con una centrifugación a 1.500 rpm durante 3 minutos y se resuspendieron en 5 mL de medio DMEM alto en glucosa (4,5 g/L) suplementado con 10% de suero fetal bovino y 0,5% de penicilina/estreptomina. Se hizo una dilución 1:10 de esta suspensión celular y se cargó en la

cámara Neubauer para estimar la densidad celular mediante el contador de células automático de Corning. El conteo reveló una densidad de $4,3 \cdot 10^6$ células/mL, por lo que se tomaron 407 μ L de esta suspensión celular ($1,75 \cdot 10^6$ células totales) y se llevaron hasta un volumen final de 14 mL con DMEM suplementado, obteniendo una disolución de $0,125 \cdot 10^6$ células/mL. A continuación, se sembraron cuatro pocillos diferentes con 2 mL por pocillo de esta suspensión, dejando 250.000 células por pocillo, y se incubaron durante 24h a 37 °C, con oxígeno a saturación y al 5% de CO₂.

Al día siguiente, se realizó la transfección celular utilizando el kit *FuGENE® HD Transfection Reagent* (Promega). Primero, se prepararon tres alícuotas con 1 μ g de plásmido silvestre a partir del stock de vector (1.825 ng/ μ L). A continuación, se preparó la mezcla de transfección por triplicado, añadiendo 94 μ L de DMEM no suplementado a 6 μ L del reactivo FuGENE. La mezcla se agitó un vórtex y se dejó incubar 5 minutos a temperatura ambiente. Se añadieron las diluciones de plásmido a la mezcla anterior y se dejó incubar de nuevo durante 15 minutos. En ese punto, se cambió el medio a las células. Finalmente, se añadió la mezcla de transfección a tres pocillos, asegurando una repartición homogénea y se devolvió el cultivo al incubador durante 48h más. Se dejó un pocillo sin transfectar como control negativo. Transcurrido el tiempo de incubación, se retiró el medio de las células y se reemplazó con tampón fosfato salino (PBS) (1X) para su observación *in vivo* mediante microscopía de fluorescencia *in vivo* (PAULA). La sustitución del medio DMEM por PBS es imprescindible, ya que el medio contiene rojo de fenol que es brillante bajo los canales de excitación del microscopio de fluorescencia, lo que limita la observación del cultivo por la difuminación de la imagen. Se visualizó el cultivo con un objetivo 10X bajo campo claro para observar la morfología celular y bajo el canal de luz verde ($\lambda_{\text{excitación}} = 480 \text{ nm}$) para detectar la presencia de la proteína de fusión EGFP-GATA2.

3.2.2.4.- Extracción de ARN y proteínas de los cultivos transfectados

Tras la observación *in vivo* al microscopio de fluorescencia, se realizó un aislamiento de ARN y proteínas de los pocillos transfectados.

Para la extracción de ARN, se empleó el kit *RNeasy® Mini Kit* (QIAGEN). Inicialmente, el cultivo celular se sometió a una disrupción enzimática y mecánica. Para ello, se eliminó el PBS de los pocillos y se realizó un lavado adicional con el mismo tampón. A continuación, se añadió tampón RLT (suplementado previamente con β -mercaptoetanol en proporción 1:1000, como inactivador de ARNasas) y se disgregó mecánicamente el cultivo con un raspador celular. Los lisados se recuperaron y se transportaron en hielo hasta una cabina de flujo para realizar la extracción de ARN por columna, siguiendo las instrucciones del fabricante. El contenido de la columna se eluyó con 35 μ L de agua libre de ARNasas. Para estimar el rendimiento de la purificación, se empleó un equipo de espectrofotometría de microvolumen NanoDrop (Thermo Fisher Scientific) para evaluar tanto la concentración de ARN como su pureza (**Tabla suplementaria 3**). Finalmente, las extracciones se preservaron a -80 °C.

Para la extracción de proteínas se siguió un protocolo optimizado propio. Para ello, se conservó el primer eluyente de la extracción de ARN, que contenía la fracción proteínica del lisado y se le

incorporaron 4 volúmenes de acetona pura por cada volumen de extracto. Esta disolución se incubó o/n a -20 °C. Al día siguiente, se preparó tampón RIPA para la extracción con 100 µL de tampón stock RIPA (10X), 143 µL de EDTA (7X) y 757 µL de agua Milli-Q. Se centrifugaron los extractos de proteína en frío (4°C, 16.000 x g, 10 minutos) y se descartó el sobrenadante. El pellet se lavó con etanol puro y se centrifugó durante 1 minuto más a la misma velocidad angular. Finalmente, se eliminó el sobrenadante, el pellet se dejó secar al aire y se resuspendió en 100 µL de tampón RIPA mediante sonicación. Las suspensiones de proteína se conservaron a -20 °C.

3.2.2.5.- Retrotranscripción y análisis de *splicing* por PCR

A partir de la extracción de ARN realizada, se obtuvo ADNc por retrotranscripción siguiendo las instrucciones del kit *PrimeScript™ RT* (Takara). Se empleó 1 µg de ARN en cada reacción, por lo que se ajustó el volumen de los reactivos para esa cantidad de ARN.

El ADNc sintetizado se amplificó por PCR empleando cebadores específicos del constructo (para validar la expresión del gen quimérico) y con cebadores en los exones 4 y 6 de GATA2 (para estudiar los patrones de *splicing* alternativo del exón 5). Se utilizó la *PCR Master Mix* (2X) (ThermoScientific) y dos parejas de cebadores diferentes: EGFP_F/Int2_R (amplicón de 877 pb) y E46_F/E46_R (amplicón de 552 pb). El programa de PCR utilizado queda esquematizado en la **Figura 4**. Puesto que la duración del paso de extensión a 72 °C depende del tamaño del amplicón a sintetizar, se utilizaron dos programas diferentes para cada pareja de cebadores.

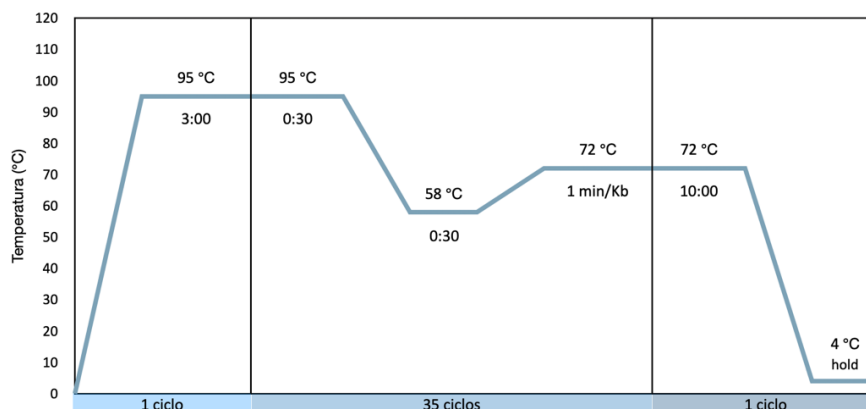


Figura 4. Programa de termociclador empleado para la amplificación del ADNc. La temperatura de hibridación de los cebadores se escogió siguiendo las instrucciones de la *Master Mix* empleada ($T_m - 5$ °C). Para los amplicones de 552 pb y 877 pb se emplearon tiempos de extensión de 30 y 55 segundos, respectivamente.

Finalmente, se reveló el resultado de las PCR mediante electroforesis en gel de agarosa (1% agarosa en tampón TBE 1X, 110 V de potencial eléctrico, durante 1h y 20 minutos), empleando el tinte fluorescente *SYBR® Safe DNA Gel Stain* (Invitrogen) como marcador de ADN y el patrón de bandas *GeneRuler 1 kb DNA Ladder* (Thermo Scientific). Alternativamente, se empleó también el equipo de electroforesis capilar QIAxcel (QIAGEN). Este instrumental es capaz de separar moléculas en función de su movilidad electroforética (relación masa-carga) en un tiempo muy reducido (12 muestras cada 6 min) y tomando volúmenes de muestra del orden de microlitros (1-10 µL).

3.2.2.6.- Limpieza del producto de PCR y secuenciación de los amplicones

Los productos de PCR obtenidos a partir de la amplificación del ADNc se secuenciaron para confirmar a nivel nucleotídico que se trataba de las regiones esperadas. Para ello, se empleó el servicio de secuenciación Sanger de STABVIDA. Sin embargo, antes de ello, los amplicones se limpiaron mediante una reacción enzimática para eliminar los componentes residuales de la PCR, que interferirían en las reacciones de secuenciación. Para ello, se utilizó el reactivo *Illustra ExoProStar* (Cytiva), que contiene fosfatasa alcalina y exonucleasa de tipo I para realizar la digestión completa de los cebadores y la inactivación de los dNTP no incorporados al ADN sintetizado, respectivamente.

3.2.2.7.- Western blot a partir del extracto de proteínas del lisado celular

Antes de comenzar con el protocolo del *Western blot*, se realizó una cuantificación del contenido proteínico de los extractos con el kit *Pierce BCA Protein Assay Kit* (Thermo Scientific). Se prepararon diluciones con BSA (*bovine serum albumin*) para generar una recta patrón. Asimismo, se diluyeron las muestras de interés en ratio 1:9 con tampón RIPA (1X). Las mediciones tanto de la curva patrón como de los extractos de interés se realizaron por duplicado (**Figura suplementaria 2**).

Para el *Western blot*, en primer lugar, se realizó un SDS-PAGE (del inglés, *sodium dodecyl sulphate - polyacrilamide gel electrophoresis*) para separar las proteínas exclusivamente en función de su masa molecular. Se preparó el *resolving gel*, se dejó polimerizar y, por encima de este, se superpuso el *stacking gel*, quedando reflejados en la **Tabla 2** los componentes de cada gel. En este punto, se prepararon alícuotas con 20 µg de proteína y tampón de carga Laemmli (1X y 2X) y se incubaron a 95 °C durante 10 minutos para provocar la desnaturalización de las proteínas. Como patrón de peso molecular, se empleó *Novex Sharp Pre-Stained* (Invitrogen).

Tabla 2. Componentes de los geles *resolving* y *stacking* del SDS-PAGE

	<i>Resolving gel</i>	<i>Stacking gel</i>
Componente (c_i)	Volumen (c_f)	Volumen (c_f)
dH₂O	3,2 mL (-)	3.075 mL (-)
Acrilamida (30%)	2,67 mL (10%)	0,625 mL (10%)
Tris-HCl (1,5M) pH 8,8	2 mL (375 mM)	1,25 mL (375 mM)
SDS (10%)	80 µL (0,1%)	50 µL (0,1%)
APS (10%)	80 µL (0,1%)	50 µL (0,15)
TEMED	8 µL (0,1%)	5 µL (0,1%)

Se presentan los volúmenes y, entre paréntesis, la concentración inicial (c_i) y final (c_f) de cada componente. El *resolving* se preparó en un volumen final de 8 mL; el *stacking* en 5 mL. Se empleó un 10% de acrilamida, óptimo para resolver proteínas de 15-100 kDa de masa molecular. *SDS*: *dodecilsulfato sódico*; *APS*: *persulfato amónico*; *TEMED*: *tetrametiletildiamina*. Los reactivos se añadieron en orden para inducir la polimerización de la acrilamida en último lugar, ya que APS y TEMED son los catalizadores de la reacción.

Tras la electroforesis, se realizó la transferencia de las proteínas del gel a una membrana comercial de fluoruro de polivinilideno (*GE Healthcare Life Sciences Amersham Hybond PVDF membrane*, Thermo Fisher Scientific) con un molde tipo sándwich. En primer lugar, se preparó el tampón de transferencia (1X) a partir de 100 mL de stock (10X), 100 mL de metanol absoluto y 800 mL de dH₂O. A continuación, se activó la membrana con metanol durante 1 minuto y se introdujo en tampón de transferencia. Se montó el molde y, para evitar el sobrecalentamiento del nuevo soporte, se introdujo la cubeta de transferencia en hielo y se corrió a un amperaje constante de 350 mA durante 1h. Pasado este tiempo, la membrana se sometió a una tinción de Ponceau para confirmar que la transferencia había funcionado.

Finalmente, la detección de proteínas se llevó a cabo mediante un ensayo de inmunodetección de tipo sándwich, empleando un anticuerpo monoclonal de conejo anti-EGFP (Invitrogen) como anticuerpo primario y un anticuerpo anti-conejo acoplado a HRP (del inglés, *horseradish peroxidase*) (anti-conejo IgG H&L, Abcam) como anticuerpo secundario. El revelado se llevó a cabo en el equipo de detección de quimioluminiscencia utilizando el kit *UltraScience Pico Ultra Western Substrate* (Bio-Helix). A posteriori, se realizó un *stripping* de la membrana y se empleó un anticuerpo monoclonal anti- β -actina-HRP (Merck) para evaluar la expresión de la β -actina como gen constitutivo.

3.3.- MUTAGÉNESIS DIRIGIDA PARA EL ESTUDIO DE VARIANTES GENÉTICAS

3.3.1.- Diseño de cebadores para mutagénesis puntual

Puesto que se empleó el kit *QuikChange Lightning Site-Directed Mutagenesis* (Agilent) para la introducción de variantes puntuales en el plásmido silvestre, se utilizó el programa recomendado por la casa comercial para el diseño de cebadores (*QuikChange Primer Design Program*, <https://www.agilent.com/store/primerDesignProgram.jsp>). En primer lugar, se introdujeron en el programa las secuencias silvestres del intrón 4-5, del exón 5 y del intrón 5-6 del gen *GATA2*. Se localizaron las posiciones nucleotídicas de las 12 variantes de interés y se especificó el cambio deseado para cada variante. Finalmente, el programa devolvió las secuencias de los cebadores directo y reverso (**Tabla suplementaria 4**) que se solicitaron al servicio de síntesis de oligonucleótidos de IDT. Los cebadores recibidos se resuspendieron en dH₂O (100 μ M) y se hicieron alícuotas de uso frecuente (10 μ M).

3.3.2.- PCR de mutagénesis y digestión del plásmido parental

Para la síntesis de plásmidos mutantes, se siguieron las instrucciones del kit comercial empleado. El método consiste en un programa de PCR optimizado (**Figura 5**) para la ADN polimerasa *QuikChange Lightning Enzyme*, de alta fidelidad de copia y alta velocidad de síntesis. Se prepararon reacciones individuales para cada pareja de cebadores (es decir, para cada variante) con 50 ng de plásmido silvestre. Tras la PCR, se obtiene una mezcla de plásmido parental y plásmido mutante. Para eliminar el plásmido silvestre, sin la variante deseada, se aprovecha la metilación diferencial de este ADN frente al plásmido recién sintetizado. La cepa bacteriana VB UltraStable (que contiene el plásmido) protege su material genético frente a la degradación por sus propias nucleasas mediante la

metilación de guaninas. Sin embargo, los plásmidos mutantes sintetizados *in vitro* carecen de esta marca epigenética, por lo que se puede emplear una nucleasa (como DpnI) que reconozca ADN metilado para seleccionar los plásmidos de nueva síntesis.

Se emplearon 10 U de la endonucleasa DpnI, que reconoce la secuencia 5'-G_mA[▼]TC-3'. La digestión del plásmido parental se realizó a 37 °C durante 10 min.

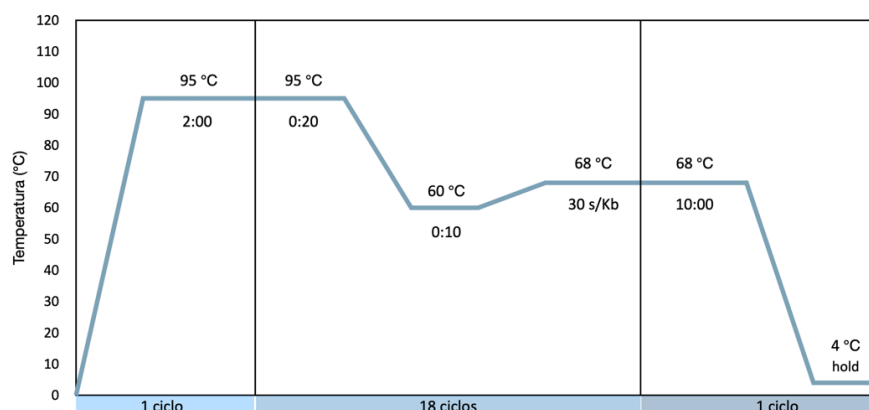


Figura 5. Programa de termociclador base para la PCR de mutagénesis. Puesto que la longitud del plásmido parental es de, aproximadamente, 11 Kb se empleó un tiempo de extensión de 5:30.

3.3.3.-Transformación bacteriana y selección en placa

Se utilizó una cepa comercial de *Escherichia coli* para la replicación de los plásmidos mutantes. En concreto, se emplearon bacterias *XL10-Gold Ultracompetent Cells* (Agilent), especialmente diseñadas para la clonación de constructos genéticos grandes.

La transformación se realizó mediante choque térmico. Para ello, previamente, las bacterias se descongelaron en hielo. Se transfirieron 2 µL del producto de mutagénesis a 50 µL de la suspensión bacteriana y se incubaron 30 minutos en hielo. Tras ello, se realizó el choque térmico con un termobloque digital a 42 °C durante 45 segundos. Las bacterias se reincubaron durante 1 minuto más en hielo y se les añadieron 450 µL de medio Luria-Bertani (LB) líquido. Finalmente, se introdujeron en un incubador a 37 °C con agitación (160 rpm) durante 2 horas. De cada cultivo resultante se sembraron dos placas de agar con ampicilina (100 µg/mL), con 50 µL y 150 µL del mismo, y estas se dejaron crecer en un incubador a 37°C o/n (durante un mínimo de 16 h).

Adicionalmente, se realizó una prueba de transformación bacteriana por electroporación tras varios intentos fallidos de obtener colonias mediante choque térmico. Para ello, se empleó una cepa bacteriana electrocompetente casera (obtenida a partir de bacterias *Escherichia coli* TOP10) y el equipo *MicroPulser Electroporator* (BioRad). Se tomaron 45 µL de las bacterias y 0,8 µL del producto de mutagénesis, que se llevaron a una cubeta de electroporación (previamente en hielo). A continuación, la cubeta se introdujo en el soporte del equipo (en modo *bacteria*) y se realizó el shock eléctrico. Finalmente, se tomó 1 mL de medio LB líquido y se resuspendió cuidadosamente la solución bacteriana. La disolución se sembró en 2 placas de agar con ampicilina (100 µg/mL). En

una placa, se incorporaron 200 μ L; se centrifugó la suspensión restante (1500 rpm, 3 min) para hacer un pellet que se disolvió en otros 200 μ L de LB líquido y se sembró la totalidad de esta disolución en otra placa.

3.3.4.- Amplificación, aislamiento y purificación de los plásmidos mutantes

Las colonias bacterianas crecidas se picaron y se disolvieron en 50 μ L de dH₂O. De esta suspensión se tomaron 25 μ L y se añadieron a 4 mL de medio LB líquido integrado con ampicilina (100 μ g/mL) para su propagación. Los cultivos se incubaron o/n a 37 °C en agitación (160 rpm). A partir de los mismos, se realizó la extracción y purificación por columna del ADN plasmídico (ADNp) siguiendo el protocolo del kit *QIAprep® Miniprep* (Qiagen). Finalmente, el ADN purificado se eluyó de la columna con 35 μ L de agua libre de nucleasas y se evaluó el rendimiento de la extracción mediante espectrofotometría con un equipo NanoDrop (Thermo Fisher Scientific)

3.3.5.- Secuenciación de la región con la variante introducida

Aunque la digestión del plásmido parental, en principio, asegura que las colonias obtenidas sean de plásmido mutante, el rendimiento de la digestión puede ser imperfecto. En este supuesto, si queda una cantidad suficiente de plásmido silvestre en el producto de mutagénesis, es posible que las bacterias transformadas no posean el vector con la variante introducida, es decir, sean falsos positivos por haber incorporado plásmido salvaje. Por este motivo, se realizó la secuenciación mediante Sanger de las regiones de *GATA2* donde es esperable encontrar las variantes para validar la presencia de estas en los clones obtenidos. Para ello, se partió de las alícuotas de plásmido purificadas y se secuenciaron con los cebadores pertinentes para cada región: las variantes en la región intrónica profunda del exón 4 se cubrieron con los cebadores I45_F y I45_R, mientras que las variantes alrededor de las posiciones canónicas de *splicing* del exón 5 se abordaron con los cebadores E5_F y E5_R.

3.3.6.- Sobreexpresión del gen quimérico mutado en células HEK293

Tras la validación de la presencia de las variantes mediante la secuenciación del vector, se transfectaron células HEK293 individualmente con cada vector mutante (1 variante/vector) con el objetivo de estudiar el efecto de las variantes a nivel transcripcional y en la expresión del gen *GATA2*. Para la transfección celular, se siguió el mismo enfoque experimental descrito en el apartado 3.2.2 (d). Sin embargo, en este caso, se partió de los plásmidos mutantes purificados a partir de los cultivos líquidos bacterianos de los clones validados. Además, un pocillo se transfectó con plásmido silvestre.

El análisis del efecto de las variantes sobre la expresión del ARN, las proteínas y el patrón de *splicing* se realizaron como se detalló en los apartados 3.2.2.4-3.2.2.7. El rendimiento y la pureza de la extracción de ARN quedan recogidos en la **tabla suplementaria 4**. Además, el resultado de la cuantificación de los extractos proteínicos se puede consultar en la **Figura suplementaria 3**.

3.4.- CUANTIFICACIÓN DE IMÁGENES Y ANÁLISIS ESTADÍSTICOS

Las observaciones cualitativas realizadas mediante microscopía de los cultivos transfectados se cuantificaron y se compararon mediante análisis estadísticos para determinar si existían diferencias significativas. De cada transfección, se tomaron 8 imágenes, capturadas en 2 momentos diferentes.

Se realizó una cuantificación de la densidad celular (área de campo claro) y de la fluorescencia (área de campo verde) de cada cultivo celular empleando el software de análisis ImageJ. Tanto las imágenes de campo claro como de campo verde se transformaron a formato de 16 bits para su tratamiento. A las primeras, se les aplicó el filtro de imagen por Varianza en un radio de 10 píxeles para reducir las diferencias de tonalidad entre el citosol y los bordes celulares. A continuación, se estableció umbral en 95 puntos con el algoritmo de detección por defecto. Para las segundas, no hizo falta aplicar un filtro de imagen, ya que el software pudo identificar correctamente las áreas verdes por defecto. Se estableció el umbral en 15 puntos con el mismo algoritmo que en las imágenes de campo claro. Finalmente, ambos tipos de imágenes se transformaron en binarias y se cuantificaron las partículas de color negro.

Los análisis estadísticos se realizaron con el software de análisis estadístico Prism (GraphPad). Se comprobó la normalidad de los datos con el test de Shapiro-Wilk. Primero, se realizó un test de correlación de Spearman con los datos de confluencia celular y fluorescencia obtenidos de los cultivos con el vector silvestre. Después, se realizó un ANOVA simple con 4 grupos (vector silvestre y 3 variantes genéticas) y un total de 32 valores (8 de cada cultivo transfectado). Finalmente, aplicamos el test de comparación múltiple de Tukey para identificar qué grupos presentaban medias significativamente diferentes. Además, también se recurrió a un test t-Student (dos colas) para realizar una comparación agrupada de los cultivos transfectados con el vector silvestre y con las variantes. Para todas las pruebas y para el cálculo de los p-valores asociados se tomó un nivel de confianza $(1-\alpha)$ del 95%.

4.- RESULTADOS

4.1.- RESULTADO DEL ANÁLISIS *IN SILICO* CON EL PREDICTOR SPLICEAI

El filtrado de las variantes genéticas con un valor de corte de SpliceAI igual o superior a 0,20 reveló un enriquecimiento de variantes que afectan al proceso de *splicing* entre el intrón 4 y 5 de GATA2. Un 77% de todas las mutaciones estaban localizadas en esta región (10 de 13 variantes), destacando aquellas que afectan a los sitios canónicos y la variante c.1055G>T (p.Cys352Phe), que podría crear un sitio críptico aceptor de *splicing*. En este punto, se priorizaron las variantes que podrían afectar a los sitios canónicos aceptor y donador de *splicing* del exón 5, seleccionando un total de 6 variantes: c.1018-2A>G, c.1018-2A>T, c.1018-G>T, c.1018-1G>A, c.1143+2T>A y c.1143+5G>C. La comparación con la base de datos de ClinVar reveló una variante no identificada previamente, de clasificación potencialmente patogénica, que afectaba al sitio de un sitio canónico aceptor en el exón 5 (c.1018-2A>C). De nuevo, se empleó la herramienta SpliceAI para predecir el impacto de esta variante en el proceso de *splicing* de GATA2. Este análisis reveló unos valores de AG y AL superiores al valor de

corte 0,20. Por lo tanto, se incluyó esta variante en la lista del estudio. Como se esperaba para las 4 variantes en la región intrónica seleccionadas y para la variante *missense*, el algoritmo no predijo ningún efecto sobre el *splicing*.

Tabla 3. Predicción del impacto de las variantes en los sitios de *splicing* del algoritmo SpliceAI

Variante	Consecuencia	AL	DL	AG	DG
c.1143+2T>A	Variante en sitio donador	0,10	1 (2 pb)	0,01	0,63 (-142 pb)
c.1143+5G>C	Variante en sitio donador	0,09	0,46 (5 pb)	0,01	0,64 (-139 pb)
c.1018-2A>T	Variante en sitio aceptor	1 (-2 pb)	0,01	0,90 (-13 pb)	0,01
c.1018-2A>G	Variante en sitio aceptor	1 (-2 pb)	0,04	0,89 (-13 pb)	0,01
c.1018-2A>C	Variante en sitio aceptor	1 (-2 pb)	0,03	0,90 (-13 pb)	0,01
c.1018-1G>T	Variante en sitio aceptor	1 (-1 pb)	0	0,91 (-12 pb)	0
c.1018-1G>A	Variante en sitio aceptor	1 (-1 pb)	0,03	0,90 (-12 pb)	0,01
c.1017+532T>A	Variante en el <i>enhancer</i> +9,5kb	0	0	0	0
c.1017+572C>T	Variante en el <i>enhancer</i> +9,5kb	0,01	0	0	0
c.1017+582G>T	Variante en el <i>enhancer</i> +9,5kb	0,01	0	0	0
c.1017+693dup	Variante en el <i>enhancer</i> +9,5kb	0	0	0,03	0,02
c.1061C>T	Cambio aminoacídico	0,01	0	0,01	0

Se presentan los valores de SpliceAI para: AG (ganancia de un sitio aceptor); AL (pérdida de un sitio aceptor); DG (ganancia de un sitio donador); DL (pérdida de un sitio donador). Entre paréntesis, se presenta la distancia hasta el sitio afectado con respecto a la variante introducida. Rojo (valores > 0,80, alta precisión), amarillo (> 0,50, precisión mínima recomendada) y verde (> 0,20, predicción pobre).

El hecho de que la mayoría de las variantes en los sitios canónicos de *splicing* de *GATA2* se sitúen en los sitios aceptor y donador del exón 5 ha motivado este estudio. En resumen, a partir del listado inicial de 149 variantes, se seleccionaron 12 para su validación funcional (**Tabla 3**).

4.2.- EFECTO DE LA TRANSFECCIÓN DE CÉLULAS HEK293 CON EL PLÁSMIDO SILVESTRE

Antes de estudiar las variantes genéticas que afectan al *splicing*, verificamos que el vector diseñado era apto para realizar los análisis *in vitro* de validación de las variantes de interés. La transfección de la línea celular HEK293 con el plásmido silvestre permitió evaluar la validez del constructo diseñado a la hora de ser reconocido por la maquinaria transcripcional celular.

En primer lugar, el ensayo de microscopía de fluorescencia *in vivo* con el equipo PAULA reveló la presencia de la EGFP en los cultivos celulares transfectados, pero no en el pocillo no transfectado (**Figura 6**), lo que constituye una prueba de la expresión específica del ORF del vector. Esta observación también indica la sobreexpresión de la proteína *GATA2* en los cultivos celulares. Asimismo, observamos una expresión abundante y homogénea de la proteína de fusión en todos los cultivos transfectados. Sin embargo, la réplica 1 (**Figura 6a**) mostró una mayor superficie de luminiscencia verde con respecto al resto de repeticiones (**Figuras 6b-g**). Cabe destacar que la réplica 1 presenta una mayor densidad celular.

Para evaluar si existía alguna correlación entre el área fluorescente y la confluencia celular realizamos una prueba de correlación de Spearman. Los datos de densidad celular y área fluorescente quedan reflejados en la **Tabla suplementaria 5**. El test confirmó una correlación elevada entre ambas variables ($r=0,89$) con un p-valor asociado <0,05 (0,033). Por lo tanto, se puede

afirmar con bastante certeza que la diferencia de fluorescencia observada en el primer cultivo se explica por la mayor densidad celular (**Figura suplementaria 4**).

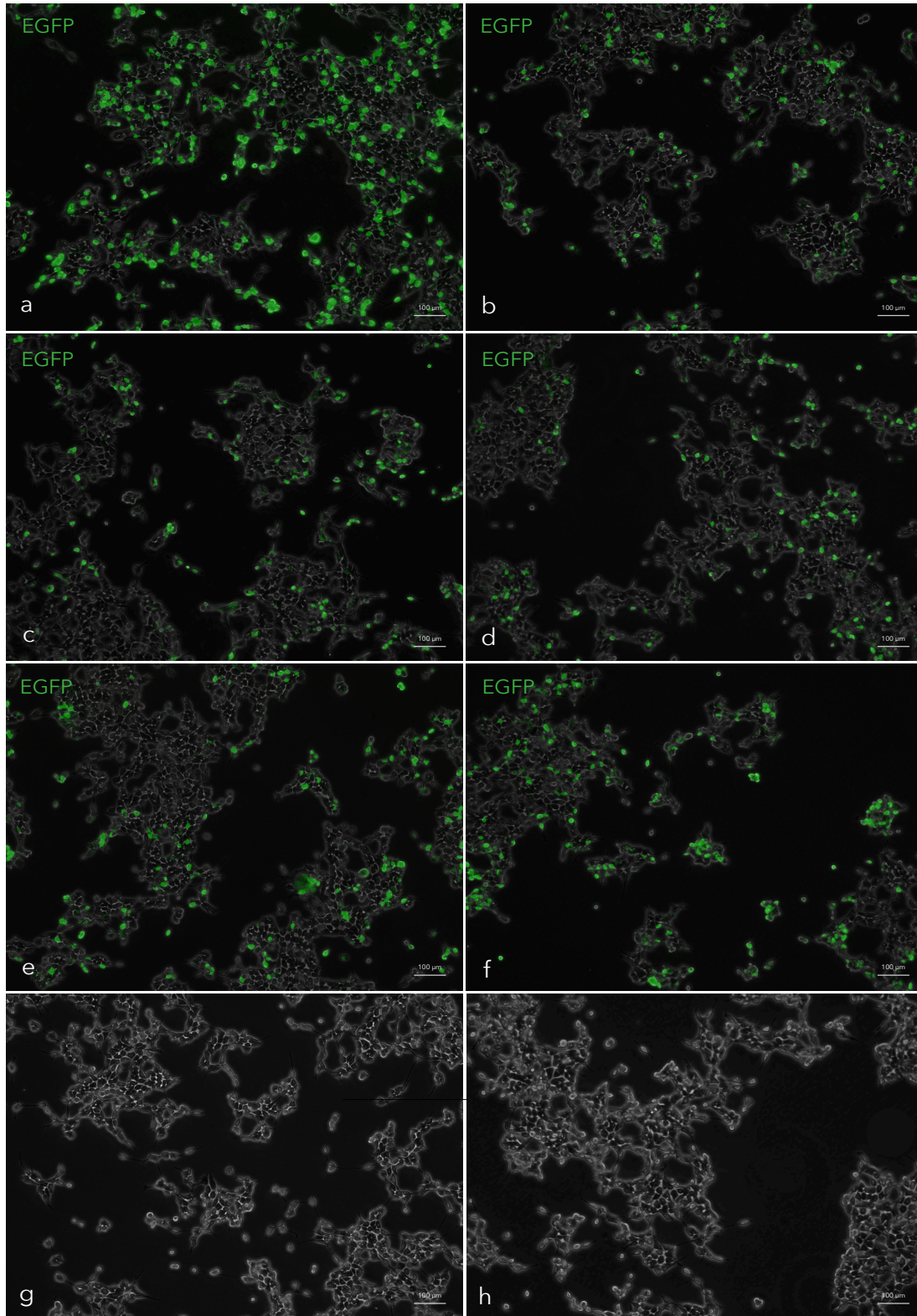


Figura 6. Imágenes de microscopía de contraste de fases y de fluorescencia de células HEK293. Objetivo 10X. LED de fluorescencia verde ($\lambda_{excitación}$: 480 nm). (a, b) Células HEK293 48 horas post-transfección con el plásmido quimérico de *GATA2* silvestre, campo claro y canal verde superpuestos (c, d) Réplica 2 (e, f) Réplica 3 (g, h) Células HEK293 no transfectadas 48h post-siembra, campo claro y canal verde superpuestos. Barra de escala: 100 μ m

La amplificación por PCR del ADNc obtenido a partir del ARN de los extractos celulares confirmó la transcripción específica del marco de lectura del vector. Se empleó un cebador directo que hibrida en la región del constructo con el ADNc de EGFP y un cebador reverso que hibrida con el exón 3 de *GATA2*, esperando un amplicón de 877 pb. La electroforesis en gel reveló amplicones únicamente en las reacciones de las células transfectadas, mientras que no se observaron amplicones en la reacción realizada con el ADNc del cultivo no transfectado (**Figura 7a**).

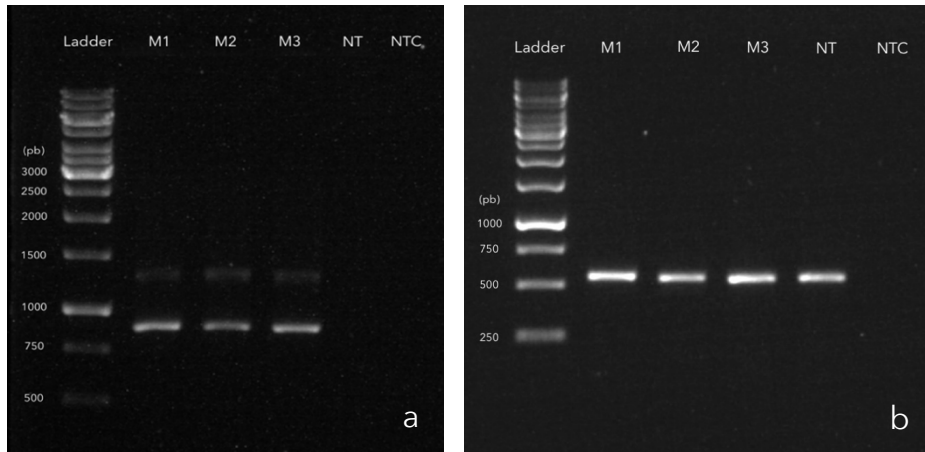


Figura 7. Resultado de la amplificación por PCR del ADNc mediante: (a) cebadores específicos del constructo y (b) cebadores específicos de *GATA2*. *Ladder*: patrón de bandas de ADN GeneRuler 1kb DNA Ladder (Thermo Scientific); *M1*, *M2*, *M3*: PCR con el ADNc obtenido a partir de la retrotranscripción de ARN de células HEK293 transfectadas con el vector (réplica 1, 2, 3); *NT*: PCR con el ADNc obtenido a partir de la retrotranscripción de ARN de células HEK293 no transfectadas. *NTC*: PCR sin ADN molde (del inglés, *no template control*).

Sin embargo, se observaron dos bandas en cada PCR, cuyo tamaño aproximado fue de 1.300pb y 800 pb, siendo más intensa esta última. El amplicón de tamaño inferior, corresponde con el amplicón esperado, ya que fue validado a nivel de secuencia mediante Sanger. La banda observada en posición superior coincide con el amplicón que se esperaba a partir del ADNp, por lo sugiere que se incorporó una cantidad residual de plásmido en la extracción de ARN. Tras la secuenciación del amplicón, las lecturas revelaron la presencia de la región del intrón 3 de *GATA2* (flanqueada por los cebadores en el ADNp), confirmando la hipótesis inicial (**Figura 8**).

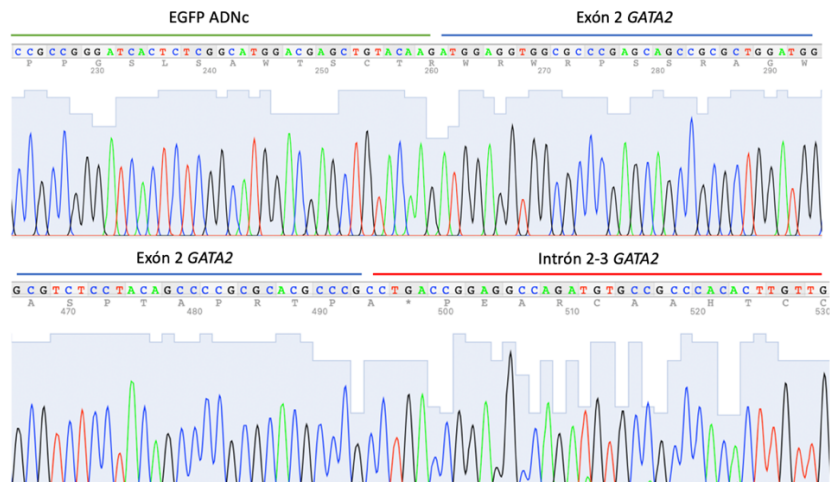


Figura 8. Resultado de la secuenciación por Sanger del ADNc de *GATA2* (Cebador EGFP_F). Se muestra la región que conecta el ADNc de EGFP y el exón 2 de *GATA2*, así como la región que muestra el intrón 3 de *GATA2*.

4.3.- EFECTO DE LA SOBREEXPRESIÓN DE GATA2 MUTADO EN CÉLULAS HEK293

Tras evaluar y comprobar que, tras la transfección del constructo genético silvestre, la proteína GATA2 se expresaba adecuadamente, se realizó la transfección de los plásmidos con las variantes seleccionadas para conocer su posible efecto. A pesar de haber realizado la mutagénesis dirigida (SDM) de todas las variantes seleccionadas (12), se lograron obtener plásmidos únicamente con las variantes c.1017+572C>T, c.1018-2A>C y c.1143+2T>A durante el tiempo experimental disponible. Por lo tanto, solo se pudo realizar el estudio de estas SNV. En apartados posteriores se presentarán las pruebas realizadas con el objetivo de optimizar la obtención de plásmido mutante y se discutirán las posibles causas del fallo repetido en la SDM.

Antes de realizar la transfección celular, la secuenciación mediante Sanger de los plásmidos purificados a partir de las colonias transformadas tras la SDM permitió validar los clones con las variantes c.1017+572C>T, c.1018-2A>C y c.1143+2T>A (**Figura 11**).

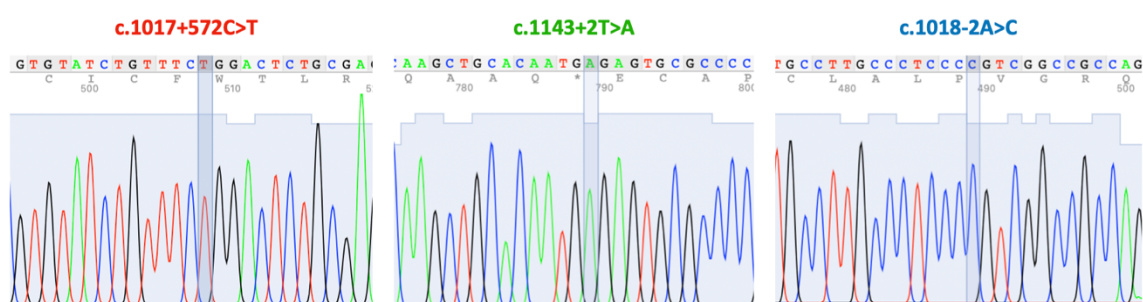


Figura 11. Electroferogramas de secuenciación mediante Sanger. Para la variante c.1017+572C>T se empleó el cebador I45_F. Mientras que para las variantes c.1143+2T>A y c.1018-2A>C se empleó el cebador E5_F.

En primer lugar, mediante microscopía *in vivo* se detectó fluorescencia verde en todos los cultivos transfectados, pero no en aquellos no transfectados (**Figura 12**). Este hecho demostró que la sobreexpresión de GATA2 era específica de los cultivos en los que se introdujo el vector. Sin embargo, a diferencia de la transfección inicial, se observaron diferencias en el nivel de fluorescencia entre el cultivo con el vector silvestre y los cultivos con el plásmido mutante. En general, el número de células que exhibieron fluorescencia verde en el cultivo transfectado con el vector silvestre (**Figura 12a**) fue superior al de los cultivos con variantes mutantes. Además, la distribución de la fluorescencia fue diferente: en las células transfectadas con el vector silvestre el campo verde era más intenso y abarcaba una mayor superficie intracelular en comparación con las células transfectadas con las variantes, como se indica con una flecha roja en la **Figura 12**.

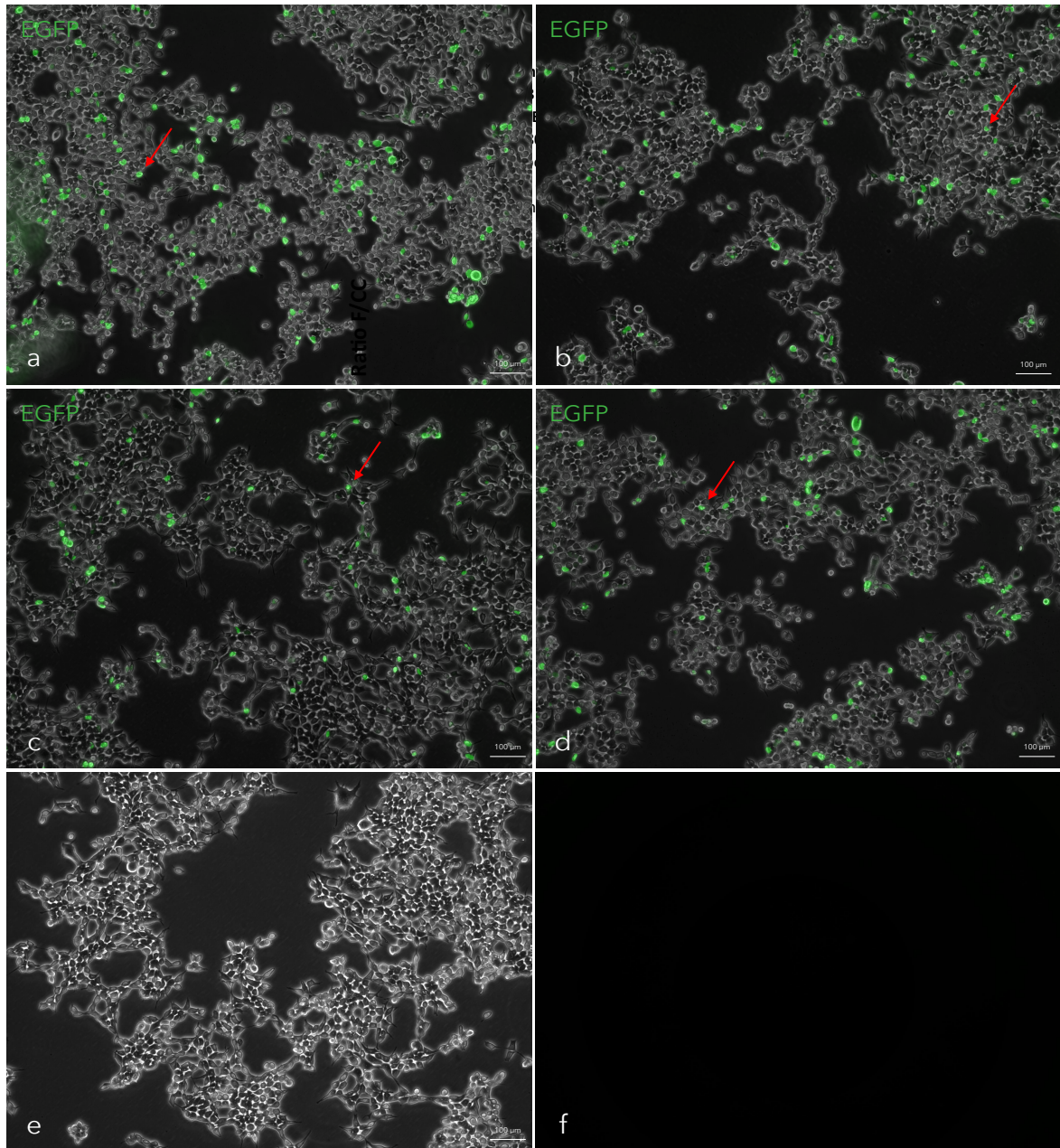


Figura 12. Imágenes de microscopía de contraste de fases y de fluorescencia de células HEK293. Objetivo 10X. Campo claro y campo verde ($\lambda_{\text{excitación}}$: 480 nm) fusionados. (a) Cultivo transfectado con el vector silvestre (WT) (b) Cultivo celular transfectado con la variante c.1017+572C>T (c) HEK293 transfectadas con la variante c.1018-2A>C (d) Células transfectadas con la variante c.1143+2T>A (e) Cultivo no transfectado (campo claro) (f) Cultivo no transfectado (campo verde). Barra de escala: 100 μm .

Para confirmar esta observación cualitativa, se realizó un análisis cuantitativo de la fluorescencia (campo verde) y de la densidad celular (campo claro) a partir de las imágenes obtenidas mediante microscopía con el software de análisis de imagen ImageJ. Puesto que anteriormente habíamos confirmado con un alto nivel de confianza la correlación entre la densidad celular y la fluorescencia, en este caso, se aprovechó esta relación para estudiar la fluorescencia y la confluencia de manera conjunta. El procedimiento completo se detalla en la **Tabla suplementaria 6**. En resumen, se calculó un cociente (ratio F/CC) entre los valores de área de campo verde y de campo claro de cada cultivo celular que devolvió el software (**Tabla 4**). De este modo, se normalizó la cuantificación de la

fluorescencia respecto a la densidad celular. A partir de estos datos, evaluamos en qué medida existían diferencias en la expresión de *GATA2* entre cada cultivo celular.

Tabla 4. Ratio F/CC calculado para cada cultivo transfectado

	WT	c.1017+572C>T	c.1018-2A>C	c.1143+2T>A
Ratio F/CC (u.a.)	0,090	0,053	0,031	0,037
	0,108	0,038	0,032	0,041
	0,131	0,052	0,028	0,026
	0,125	0,034	0,025	0,030
	0,070	0,035	0,031	0,028
	0,064	0,045	0,028	0,026
	0,071	0,044	0,024	0,030
	0,086	0,036	0,021	0,038

Para evaluar la similitud en la expresión de *GATA2* entre los cultivos transfectados, realizamos un ANOVA simple. Este análisis reveló que las medias de los grupos eran significativamente diferentes ($F=38,61$, $p\text{-valor}<0,0001$, $R^2=0,81$). Después, se realizó una comparación por grupos con el test de comparación múltiple de Tukey (**Figura 13**). Esta prueba mostró un nivel de significancia elevado al comparar la expresión de *GATA2* de cada cultivo transfectado con las variantes con la expresión de *GATA2* del plásmido silvestre (**Figura 13a**). Sin embargo, no se encontraron diferencias significativas ($p\text{-valor}>0,05$) que indicaran que la expresión de *GATA2* era diferente entre los cultivos con las variantes c.1017+572C>T, c.1018-2A>C y c.1143+2T>A (**Figura 13b**)

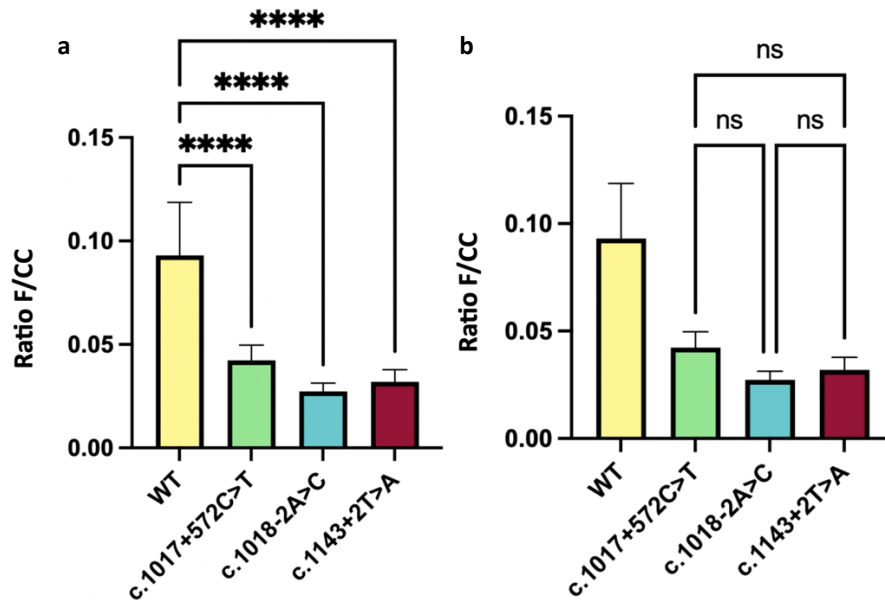


Figura 13. Comparativa de los valores de área de campo verde y de campo claro (ratios F/CC) de los cultivos celulares transfectados. Se presenta la media de los ratios, así como la barra de desviación típica. WT: cultivo celular transfectado con el plásmido silvestre (n=8); c.1017+572C>T, c.1018-2A>C y c.1143+2T>A: cultivo celular en que se ha sobreexpresado el SNV etiquetado (n=8). **** (p-valor < 0,0001). ns: no significativo (p-valor > 0,05).

Una vez comprobado que no había diferencias estadísticamente significativas en cuanto a la expresión de *GATA2* entre los cultivos transfectados con diferentes plásmidos mutantes,

procedimos a evaluar si había diferencias significativas entre los cultivos transfectados con variantes (considerándolos como un único grupo) y el cultivo con el vector silvestre. Para este análisis, se realizó un test t con los valores F/CC de ambos grupos. El resultado mostró diferencias con un alto nivel de significancia (p-valor < 0,0001) entre el grupo control y el grupo de cultivos transfectados (**Figura 14**). Por ello, rechazamos la hipótesis nula y concluimos que la inclusión de las variantes seleccionadas afecta negativamente a la expresión de *GATA2*.

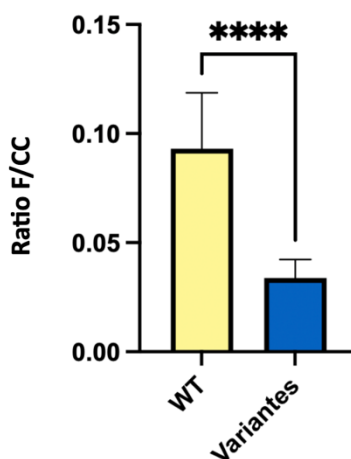


Figura 14. Comparativa de los valores de área de campo verde y de campo claro (ratios F/CC) de las variantes estudiadas agrupadas. Se presenta la media de los ratios calculados, así como la barra de desviación típica. **** (p-valor < 0,0001). WT: cultivo celular transfectado con el plásmido silvestre (n=8); Variantes: conjunto de los cultivos celulares transfectados con las SNV c.1017+572C>T, c.1018-2A>C y c.1143+2T>A (n=24).

En segundo lugar, mediante PCR, se amplificó el ADNc obtenido por retrotranscripción del ARN extraído de los cultivos celulares. Se emplearon dos parejas de cebadores: EGFP_F/Int2_R, como cebadores específicos del vector; y E46_F/E46_R, que hibridan en los exones 4 y 6 de *GATA2*, respectivamente. La primera pareja sirvió para demostrar la expresión específica del ORF del plásmido en los cultivos celulares, mientras que la segunda pareja era informativa del patrón de *splicing* alrededor del exón 5. Este análisis fue útil para valorar el posible impacto de la inclusión de las variantes en el reconocimiento de los sitios canónicos aceptores y donadores de *splicing* situados alrededor de este exón.

Cuando estudiamos las reacciones específicas del vector (**Figura suplementaria 5**), se esperaba encontrar un pico de 877 pb, correspondiente a la región cubierta por los cebadores EGFP_F y EGFP_R sobre el ADNc del vector (**Figura 15b**). Estos picos eran previsible únicamente sobre las reacciones que corresponden a las extracciones de ARN de los cultivos celulares transfectados, pues expresan el ORF del plásmido. Los electroferogramas confirmaron la existencia de amplicones de alrededor de 800 pb en las reacciones silvestre, c.1017+572C>T, c.1018-2A>C, y c.1143+2T>A, pero no en la reacción del cultivo no transfectado. Estos resultados ratificaron la tendencia esperada: solo en los cultivos transfectados hubo amplificación con los cebadores EGFP_F y Int2_R. Sin embargo, los electroferogramas mostraron un pico inesperado de 1250 pb en estas reacciones, correspondiente al el tamaño esperado si el molde fuese ADNp (**Figura 15c**). Por lo tanto, es probable que durante la extracción de ARN se produjera el arrastre de vector.

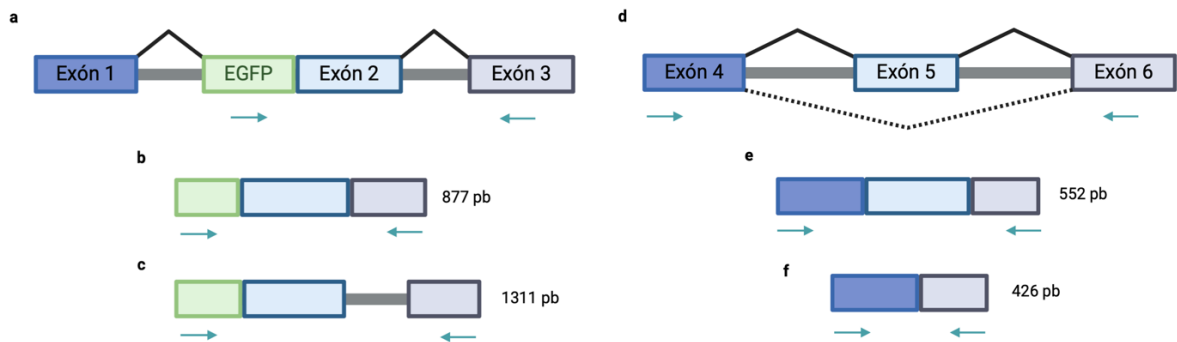


Figura 15. Esquema del patrón de *splicing* silvestre de *GATA2*. **a** Región del gen quimérico desde el exón 1 hasta el exón 3 de *GATA2* **b** Amplicón generado con los cebadores EGFP_F/Int2_R sobre ADNc **c** Amplicón generado con los cebadores EGFP_F/Int2_R sobre ADNp. **d** Región del gen quimérico desde el exón 4 hasta el exón 6 de *GATA2* **e** Amplicón generado con los cebadores E46_F/E46_R sobre ADNc **f** Amplicón generado tras el salto del exón 5. Los cebadores se representan como flechas verdes. Imagen: Created with BioRender.com

En los electroferogramas de las PCR realizadas con los cebadores E46_F y E46_R se observó un pico a alrededor de 500 pb sobre el ADNc de todos los cultivos estudiados, excepto para la variante c.1143+2T>A, que fue de tamaño inferior (400 pb) (**Figura 16b**).

Analizando cada caso por separado, en primer lugar, era esperable que el cultivo no transfectado resultara positivo para esta PCR, ya que la línea celular HEK293 expresa de manera natural el gen *GATA2*. Sin embargo, la intensidad de fluorescencia del pico de 492 pb no superó las 1.500 unidades de fluorescencia relativa (RFU). Aunque no se puede realizar una cuantificación precisa mediante una PCR convencional, si comparamos la intensidad de la fluorescencia del pico de 492 pb entre el cultivo no transfectado (**Figura 16b₂**) y el resto, podemos hacer una estimación de la cantidad de proteína *GATA2* que se expresa de manera basal en las células HEK293. Comparando este pico con el pico del cultivo transfectado silvestre (que supera las 3500 RFU), la diferencia de expresión de *GATA2* fue aproximadamente 3 veces menor en las células HEK293 no transfectadas. Por su parte, se observó un conjunto de picos a 418, 626 y 1754 pb en el cultivo con la variante c.1018-2A>C (**Figura 16b₄**) que corresponde con patrón de *splicing* generado por la propia variante. Estas observaciones confirman que las PCR fueron específicas y que las diferencias en los patrones de *splicing* y expresión de *GATA2* pueden atribuirse a las variantes introducidas en el experimento.

A nivel nucleotídico, la secuenciación mediante Sanger de los amplicones permitió comprobar, en primer lugar, que la región amplificada con los cebadores EGFP_F/Int2_R correspondía con la región del vector flanqueada por dichos oligonucleótidos. Esto ratificó la expresión específica del ORF del plásmido en todos los cultivos transfectados, pero no en el cultivo de células HEK293 sin transfectar (**Figura suplementaria 6**). Asimismo, se confirmó la presencia de ADNp residual en el extracto de ARN, ya que en las lecturas se pudo identificar la región del intrón 2 de *GATA2*. En segundo lugar, la secuenciación de los amplicones generados con los cebadores E46_F y E46_R reveló diferencias en el patrón de *splicing* de los transcritos extraídos a partir de los cultivos celulares (**Figura suplementaria 7**). La secuencia nucleotídica de los cultivos no transfectado, transfectados con el vector silvestre y transfectado con la variante c.1017+572C>T fue idéntica, correspondiendo a la secuencia contigua de los exones 4, 5 y 6 de *GATA2*. Sin embargo, la secuenciación del ADNc del

cultivo con el SNV c.1143+2T>A reveló la secuencia continua de los exones 4 y 6, lo que indica el salto del exón 5 completo.

Finalmente, como se aprecia en la electroforesis capilar y en la imagen del gel, la introducción de la variante c.1018-2A>C se reflejó en un *splicing* complejo, resultante en un conjunto heterogéneo de amplicones obtenidos con los cebadores E46_F y E46_R. Por ello, la secuenciación por Sanger del producto de PCR no pudo identificar con exactitud la extensión completa de la secuencia debido al solapamiento de los picos en el electroferograma. Sin embargo, las lecturas permitieron observar la secuencia íntegra de los exones 4 y 6 de *GATA2* y, entre ellos, el exón 5 de *GATA2* sin los 11 pb de su extremo 5'.

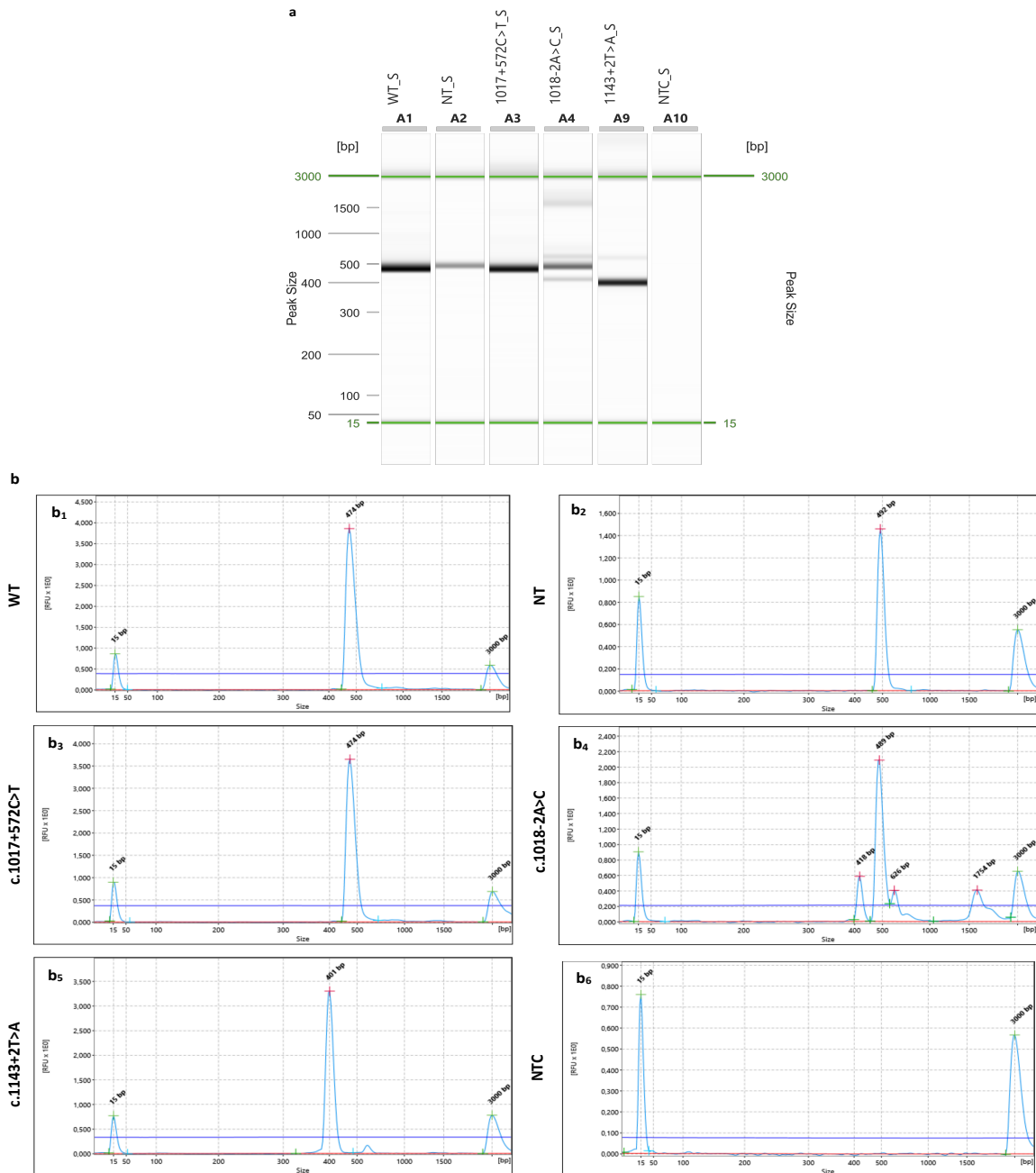


Figura 16. Amplificación por PCR con los cebadores E46_F/E46_R de ADNc. a Representación del frente de avance de la electroforesis capilar. WT (plásmido silvestre), NT (cultivo celular no transfectado), variantes c.1017+572C>T, c.1018-2A>C, c.1143+2T>A y NTC: *no template control* **b Electroferogramas individuales con las amplificaciones por PCR.** Límite inferior: 15 pb. Límite superior: 3000 pb. El umbral de RFU se ajustó a mano para mejorar la visualización de los resultados.

A nivel de proteína, el *Western blot* realizado con los extractos proteínicos de los cultivos celulares reveló bandas para todos los cultivos excepto para el cultivo no transfectado (**Figura 17**). Esto confirmó la tendencia esperada, ya que el paratopo del anticuerpo primario era específico de la proteína verde fluorescente, permitiendo la detección de la proteína de fusión EGFP-GATA2.

Se observaron diferencias en dos niveles entre los cultivos transfectados con los plásmidos que contenían las variantes y con el vector silvestre. Por un lado, el número y la posición de las bandas fueron aspectos diferenciales. En silvestre y en la variante c.1017+572C>T solo se detectó una banda que corrió hasta la posición de 80 kDa, con la variante c.1143+2T>A el tamaño de la banda fue ligeramente inferior y, por último, en la variante c.1018-2A>C se detectaron tres bandas independientes de tamaños entre 60 y 80 kDa.

Por otro lado, la intensidad de las bandas, que es proporcional a la cantidad de proteína, difirió entre las muestras. El extracto silvestre mostró la máxima intensidad de entre todos los extractos. La cantidad de la proteína de fusión en los extractos de las variantes c.1017+572C>T y c.1143+2T>A fue similar, mientras que las bandas observadas en el pocillo de la variante c.1018-2A>C resultaron de baja intensidad.

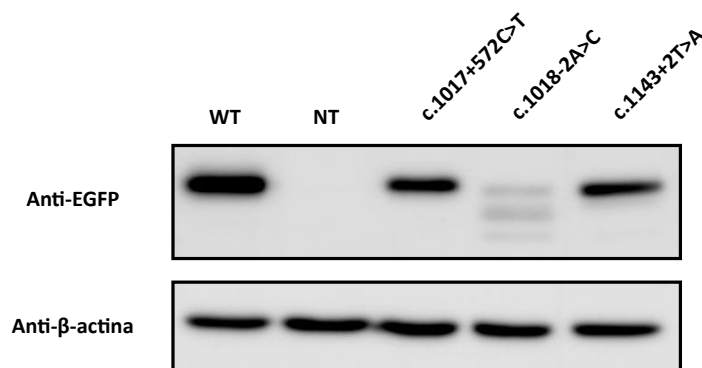


Figura 17. *Western blot* con los extractos de proteína de los cultivos celulares. WT: extracto proteínico de células HEK293 transfectadas con el plásmido pRP[Exp]-CMV>{GATA2+EGFP}; NT: extracto proteínico de células HEK293 no transfectadas. Las bandas superiores corresponden a la proteína de fusión EGFP-GATA2, mientras que las bandas inferiores corresponden a la proteína β -actina, de 42 kDa. SDS-PAGE (10% de acrilamida).

Para estimar la masa de cada una de las bandas observadas, se realizó una regresión lineal que comparaba el logaritmo de la masa molecular de cada banda con la distancia recorrida (**Figura suplementaria 8**). Esto se explica ya que la posición relativa de las bandas sigue una tendencia logarítmica respecto a la masa molecular de las proteínas. Los tamaños resultantes de esta estimación fueron para los extractos silvestre, c.1017+572C>T, c.1018-2A>C y c.1143+2T>A: 77,7 kDa; 77,2 kDa; 73,2-68,0-64,0 kDa; y 73,2 kDa, respectivamente.

Finalmente, con el objetivo de identificar cada una de las bandas que resultan del *splicing* complejo asociado a la variante c.1018-2A>C, realizamos una predicción *in silico* con los eventos identificados a nivel de ARN empleando la herramienta de traducción ExPASy. En primer lugar, en la secuencia de ADNc de *GATA2* se introdujo la pérdida de 11 pb que habíamos observado a nivel nucleotídico en el extremo 5' del exón 5. Considerando la masa de la EGFP, la herramienta predijo una proteína

truncada de 68,48 kDa. En segundo lugar, estudiamos la posibilidad de que la variante indujera una retención intrónica parcial. Para ello, antes de evaluar el efecto en la traducción de los transcritos, utilizamos la versión 0.9 del predictor NNSPLICE (https://www.fruitfly.org/seq_tools/splice.html) con el objetivo de identificar posibles sitios aceptores de *splicing* en el intrón 4 de *GATA2* que pudieran haberse activado como consecuencia de la variante. En el predictor Expasy se estudió el impacto que los diferentes sitios crípticos de *splicing* en la región pudieran tener sobre la secuencia primaria de la proteína. Entre todos, destacamos el primer evento identificado (con una puntuación de 0,95) asociado a la creación de un sitio aceptor a 1214 pb de distancia del sitio canónico de *splicing*, que induciría una proteína truncada de 64,96 kDa. El **anexo 2** recoge el análisis conjunto realizado con ambas herramientas.

5.- DISCUSIÓN

5.1.- USO DEL VECTOR pRP[Exp]-CMV>{GATA2+EGFP} PARA EL ESTUDIO DE VARIANTES GENÉTICAS EN LAS REGIONES INTRÓNICAS DE *GATA2*

Aunque las herramientas bioinformáticas actuales han mejorado notablemente en la predicción del impacto que ciertas variantes genéticas tienen sobre el proceso del *splicing*, estos programas no pueden utilizarse como herramientas de diagnóstico por sí solos. Además, tampoco permiten conocer con exactitud el impacto que estas predicciones tienen sobre los mecanismos biológicos que subyacen a la patogenicidad de las alteraciones genéticas (Desviat et al., 2012). Por ello, normalmente, para el estudio experimental de variantes que se espera tengan un efecto potencial sobre el proceso de *splicing*, se suele recurrir a los ensayos de minigenes. Estos ensayos permiten realizar una validación individualizada que otras técnicas, como la secuenciación de ARN (RNAseq) o la retrotranscripción y amplificación por PCR (RT-PCR) no pueden ofrecer (Putscher et al., 2021).

Los minigenes no son más que vectores plasmídicos con 2 exones que delimitan una región intrónica con un sitio de clonaje múltiple que permite la inserción de un fragmento genómico de interés de hasta un par de kilobases. Este fragmento suele contener la región de interés afectada por el evento mutacional y cierta extensión de las regiones flanqueantes. La transfección celular de este plásmido, que puede actuar como si fuera un gen, permite el estudio del patrón de *splicing* mediante una extracción de ARN, facilitando la comparación entre los transcritos producidos por el constructo silvestre y por el mutado (Gaildrat et al., 2010).

Entre los eventos observados, los transcritos sintetizados a partir de la secuencia mutada pueden inducir un salto del exón, por ejemplo, mientras que la versión silvestre suele incluir el exón de estudio (Desviat et al., 2012). Sin embargo, los ensayos de minigenes no permiten estudiar una variante en su contexto génico completo. Solo en los casos donde se han diseñado específicamente para ello, los minigenes pueden inducir la síntesis de proteínas si se incluyen las secuencias genéticas necesarias.

Por ello, en este estudio hemos diseñado un plásmido que incorpora la mayor parte de la secuencia genómica del gen *GATA2*, considerando el papel clave de los intrones en el estudio funcional de un

gen (Naberezhnov et al., 2021). En el caso de *GATA2*, por ejemplo, la inclusión de las regiones intrónicas es crucial, ya que el intrón 4 contiene la región *enhancer* +9,5 kb, fundamental para la hematopoyesis. En esta región se ha descrito la variante patogénica c.1017+572C>T, una de las más frecuentes que causan G2DS. El diseño del ORF del vector nos ha permitido estudiar el efecto de esta variante tanto a nivel transcripcional, como a nivel proteico, algo que no habría sido posible con un ensayo de minigenes debido a su naturaleza. El éxito de esta estrategia reside en el hecho de que la inclusión de los intrones quiméricos no altera la síntesis de la proteína *GATA2*, ya que el espliceosoma celular es capaz de reconocer los sitios canónicos de *splicing*, produciendo una proteína con la masa molecular esperada. Además, la secuenciación por Sanger del ADNc de los cultivos celulares corroboró que el patrón de *splicing* de los transcritos con el vector silvestre y los del cultivo celular no transfectado son idénticos. Igualmente, el *Western blot* de los extractos proteicos confirmó la presencia de una proteína de aproximadamente 80 kDa, que corresponde a la proteína de fusión EGFP-*GATA2*, con masas individuales de 26,9 kDa y 50,5 kDa.

5.2.- INTRODUCCIÓN DE VARIANTES GENÉTICAS MEDIANTE PCR DE MUTAGÉNESIS DIRIGIDA DE SITIO CON CEBADORES SOLAPANTES

La introducción de las variantes genéticas sobre la secuencia de *GATA2* en el vector de expresión es un paso fundamental tras comprobar que el plásmido es reconocido por el cultivo celular. En nuestro caso, sin embargo, este paso ha sido limitante. El método que empleamos se basa en el uso de cebadores solapantes que contienen el cambio nucleotídico en el centro del oligonucleótido y que, mediante un programa de PCR, son capaces de sintetizar moléculas de vector mutante a partir del vector original. Hasta la obtención de plásmidos con variantes, realizamos 5 pruebas diferentes sobre el protocolo.

En primer lugar, testamos las condiciones del protocolo con los cebadores para las variantes c.1143+2T>A y c.1143+5G>C. Para ello, empleamos las condiciones estándar del kit (*QuikChange Lightning Site-Directed Mutagenesis*): 50 ng de ADN molde y el programa de PCR descrito, ajustado a los 11 kb del plásmido (5:30 minutos de extensión). Sin embargo, tras la transformación bacteriana con 2 µL de producto de mutagénesis no creció ninguna colonia. Tras este primer intento, repetimos la misma aproximación para una variante no testada (c.1061C>T) y para c.1143+5G>C, con el objetivo de asegurarnos de que no habíamos cometido un error experimental, pero la SDM fue infructuosa de nuevo. A continuación, probamos a ampliar el tiempo de extensión 30 segundos por encima del tiempo estipulado por el protocolo y repetimos la PCR de mutagénesis para una variante nueva (c.1017+532T>A) y para el cambio c.1061C>T. De nuevo, realizamos la transformación por choque térmico de las bacterias competentes XL10-Gold, pero no obtuvimos ninguna colonia. En este punto, nos planteamos si el paso que resultaba limitante era la PCR de mutagénesis y, para testar esta hipótesis, intentamos la transformación bacteriana con 100 ng de plásmido silvestre. En esta ocasión, sí obtuvimos clones, aunque la eficiencia de transformación fue más baja de lo que se esperaba, considerando la elevada cantidad de plásmido empleada para la transformación. Por ello, concluimos que tanto la PCR como la transformación por choque térmico podrían estar fallando para la obtención de plásmidos con las mutaciones puntuales. Retomamos la SDM con variantes no

testadas hasta el momento (c.1017+693dup y c.1018-1G>A) y, en esta ocasión, descendimos 2 °C la temperatura de hibridación del programa de PCR (58 °C) con el objetivo de hacer menos restrictivas las condiciones de unión de los cebadores al vector molde. Asimismo, empleamos cantidades crecientes de plásmido (25, 50 y 100 ng) para tratar de maximizar la cantidad de plásmido mutante sintetizado. La transformación por choque térmico se realizó exactamente a 42 °C durante 45 segundos, como especifica el protocolo comercial de manera estricta. Nuevamente, no hubo crecimiento de bacterias en las placas. Ante esta situación, aunque el kit *QuikChange Lightning Site-Directed Mutagenesis* (Agilent) está optimizado para la introducción de cambios en plásmidos de entre 4-14 kb según el fabricante, decidimos cambiar de kit comercial para explorar la posibilidad de que el tamaño de nuestro vector (11 kb) fuese incompatible con dicho método. Así, utilizamos el kit *QuikChange XL Site-Directed Mutagenesis* de la misma casa comercial, que está específicamente diseñado para la introducción de mutaciones en vectores de entre 8-14 kb. Se realizó la SDM de las variantes c.1061C>T y c.1143+5G>C, previamente testadas, empleando 10 y 40 ng de plásmido. Sin embargo, el cambio del kit comercial tampoco sirvió para la obtención de vectores mutantes. En este punto, observamos que el porcentaje de GC de la región donde buscábamos introducir las variantes era elevado (60%), por lo que realizamos una última prueba con las mismas variantes empleando DMSO (5%) con el objetivo de disminuir la formación de estructuras secundarias en los cebadores y en el ADN molde. Nuevamente, la SDM resultó ineficaz. En último lugar, intentamos la transformación bacteriana mediante electroporación con el objetivo de evaluar si el obstáculo principal había sido la transformación mediante choque térmico, empleando los productos de mutagénesis de la variante c.1061C>T obtenidos previamente con ambos kits (Lightning y XL). Esta última prueba tampoco fue exitosa. Tras este conjunto de intentos sin éxito, decidimos retomar el protocolo inicial y repetir la SDM de todas las variantes, incluyendo aquellas no testadas hasta la fecha. Para nuestra sorpresa, esta nueva ronda de mutagénesis nos permitió obtener vectores mutantes con 3 variantes diferentes: c.1017+572C>T y c.1018-2A>C (testadas por primera vez) y c.1143+2T>A, que habíamos intentado en una ocasión.

Las causas de la imposibilidad continuada de obtener clones con ciertas variantes, así como del éxito del protocolo para otras son complejas de analizar. Por un lado, algunos autores proponen que el sistema de mutagénesis mediante PCR con cebadores solapantes posee limitaciones de partida. Zhang et al. (2021) exponen que la eficiencia del método desarrollado por Stratagene (ahora Agilent) es muy baja por la formación de dímeros de cebadores, que llevan al fallo de la reacción de PCR desde el inicio. Asimismo, si se completa la primera ronda de PCR, el producto contiene una rotura de cadena simple que impide que pueda ser utilizado como molde en el siguiente ciclo (Liu et al., 2008). Por ello, las limitaciones técnicas del propio protocolo no son despreciables. De hecho, múltiples autores han desarrollado nuevas estrategias basadas en el método de PCR empleando cebadores parcialmente solapantes, con resultados satisfactorios (Liu et al., 2008; Wang et al., 2011; Zheng et al., 2004). Sin embargo, a nivel biológico, también es cierto que el estudio de regiones genéticas complejas (con un alto porcentaje de ADN repetitivo, una proporción elevada GC o que formen estructuras secundarias) puede subyacer la bajísima eficiencia de mutagénesis que hemos observado. Con todo ello, no es descartable la necesidad de abordar la introducción de mutaciones

empleando otros sistemas, como CRISPR/Cas9, por no ser un sistema basado en PCR que ha demostrado una eficiencia y precisión sin precedentes (Hryhorowicz et al., 2017).

5.3.- ESTUDIO DEL EFECTO DE LAS VARIANTES GENÉTICAS SOBRE LA EXPRESIÓN DEL GEN *GATA2* EN CÉLULAS HEK293

La transfección celular con los plásmidos que albergan las variantes c.1017+572C>T, c.1018-2A>C y c.1143+2T>A nos ha permitido evaluar el efecto de las mismas sobre la expresión de *GATA2*. Cabe diferenciar entre la primera variante, localizada en la región intrónica profunda del intrón 4 de *GATA2*, y las dos últimas variantes, que afectan a los sitios canónicos aceptor y donador, respectivamente, del exón 5.

La variante c.1017+572C>T produce una transición nucleotídica en una región esencial reguladora de la transcripción de *GATA2*: el *enhancer* +9,5kb. Este cambio de nucleótido afecta al motivo de unión (TTCC) reconocido por la superfamilia de factores de transcripción Ets para promover la síntesis de transcritos de este gen (Churpek et al., 2019). Se ha sugerido que esta mutación puntual disminuye significativamente la expresión de *GATA2* (Hsu et al., 2013), como queda confirmado por nuestra validación. A nivel nucleotídico, los transcritos maduros no presentan alteraciones, pudiéndose apreciar íntegros los exones 4, 5 y 6 de *GATA2*. Sin embargo, a nivel proteico, se observa una pérdida significativa de expresión en comparación con el cultivo transfectado con el vector que contiene la secuencia silvestre de *GATA2*. Clínicamente, esta variante es una de las más reportadas en los pacientes afectados por G2DS y está clasificada como patogénica con los criterios ACMG/AMP PS3, PS4, PM1, PM2 y PP1_Strong (Homan et al., 2021). No obstante, muy pocos estudios se han centrado en la caracterización funcional de esta variante, especialmente en muestras humanas (Hsu et al., 2013; Soukup et al., 2019). Por ello, la evidencia de la pérdida de expresión que observamos en este estudio es muy informativa para explicar la haploinsuficiencia de *GATA2* asociada a la presencia de esta variante en heterocigosis (Spinner et al., 2014).

La variante c.1018-2A>C se localiza en la región intrónica aguas arriba del exón 5 y afecta al sitio aceptor de *splicing* de dicho exón. Lógicamente, al ser un cambio nucleotídico que afecta al motivo de reconocimiento de la proteína U2AF1 del espliceosoma, este se va a traducir en una alteración del patrón normal de *splicing* por la pérdida o disminución en el reconocimiento del sitio canónico (Bonnal et al., 2020). A nivel *in silico*, la herramienta de *deep-learning* SpliceAI predice la pérdida del sitio aceptor y el reconocimiento de un sitio aceptor críptico a una distancia de 13 pares de bases aguas abajo de la posición de la variante, en la región codificante del exón 5. Nuestro estudio ha identificado un transcrito en el que las primeras 11 pb en el extremo 5' del exón 5 están delecionadas mediante la secuenciación del ADNc del cultivo celular transfectado con el plásmido que contiene la variante c.1018-2A>C. Esta pérdida podría inducir el desplazamiento del marco de lectura a partir de este punto. Además, la amplificación por PCR de la región afectada y el posterior revelado mediante electroforesis capilar exhiben amplicones consistentes con la retención del intrón 4, así como del salto del exón 5. A nivel proteico, el *Western blot* revela 3 bandas de tamaño inferior (hasta 20 kDa de diferencia) a la proteína silvestre. La banda superior podría corresponder

al salto del exón 5 de *GATA2*, mientras que las bandas inferiores podrían ser acordes a una retención intrónica (que produce un codón de parada prematuro e induce la síntesis de una proteína truncada), así como a un posible desplazamiento del marco de lectura a causa del sitio críptico creado en el exón 5. A nivel bioinformático, estas hipótesis cobran fuerza, ya que la herramienta de traducción ExPasy predice tamaños de proteína muy similares a los observados mediante *Western blot* al introducir estas alteraciones en la secuencia silvestre de *GATA2*. Considerando la creación de un sitio críptico aceptor en el exón 5, que elimina 11 pb en el extremo 5' de este exón, esta herramienta predice una proteína de 68,48 kDa. Además, si asumimos una retención intrónica parcial, el predictor estima que existe la posibilidad de que se produzca una proteína de 64,96 kDa. Por lo tanto, existe un nivel de evidencia adicional respecto al patrón de *splicing* complejo asociado a la variante c.1018-2A>C. Sin embargo, harían falta estudios a nivel proteómico para confirmar esta hipótesis. A nivel clínico, esta variante se ha reportado como potencialmente patogénica, pero su estado de revisión es limitado. Nuestro estudio funcional que presentamos acerca de la variante c.1018-2A>C podría ayudar a actualizar el criterio PVS1_MODERATE a PVS1_VERYSTRONG (RNA) (Walker et al., 2023), lo que pone de manifiesto la importancia de los estudios funcionales.

Finalmente, el SNV c.1143+2T>A se localiza aguas abajo del exón 5 de *GATA2* y constituye una variante de sitio canónico donador de *splicing*. Del mismo modo que la proteína U2AF1 reconoce el sitio 3' aceptor de *splicing*, la ribonucleoproteína U1 detecta la región 5' donadora de *splicing* (Bonnal et al., 2020). Precisamente, esta variante produce una transversión nucleotídica en el sitio donador de *splicing* aguas abajo del exón 5 de *GATA2*. SpliceAI predice una pérdida del sitio canónico donador y una ganancia de un sitio críptico 142 pb aguas arriba. La incorporación de 142 nucleótidos adicionales podría crear un codón de parada prematuro. Nuestro estudio ha revelado que este cambio nucleotídico induce el salto completo del exón 5 de *GATA2*, consistente con un estudio previo realizado mediante un ensayo de minigenes por Oleaga-Quintas et al. (2021). En nuestro caso, también hemos podido ratificar la pérdida del exón 5 a nivel proteínico (en pauta), pues el *Western blot* realizado con el extracto de proteínas del cultivo celular transfectado con el vector mutante para esta variante revela una pérdida de masa molecular de aproximadamente 4,5 kDa, correspondiente a los 42 aminoácidos perdidos del exón 5 (considerando que cada aminoácido posee una masa molecular promedio de 0,11 kDa). El salto del exón 5 implica la pérdida de la región que codifica para el dominio ZF2 de la proteína *GATA2* (Kotmayer et al., 2022), que media la interacción de este factor de transcripción con el ADN y le permite controlar la transcripción de sus genes diana (Hirabayashi et al., 2017; Rajput et al., 2023; Leubolt et al., 2020). Por lo tanto, es posible que esta variante genere consecuencias fenotípicas similares a las variantes de tipo *missense* o *nonsense* que afectan el exón 5 de *GATA2*.

En resumen, el estudio funcional de las variantes c.1017+572C>T, c.1018-2A>C y c.1143+2T>A ha ampliado nuestro conocimiento sobre el impacto a nivel molecular de estos SNV, subrayando la importancia de la caracterización funcional para mejorar la caracterización de síndromes genéticos, como la deficiencia en *GATA2*.

6.- CONCLUSIONES

- 1.- El diseño de un vector plasmídico que incluye la secuencia genómica completa del gen *GATA2* fusionado con el ADNc de EGFP y algunos intrones quiméricos para reducir el tamaño final del vector, no altera la expresión de *GATA2* ni la producción de la proteína silvestre.
- 2.- El vector diseñado permite el estudio de variantes genéticas localizadas en regiones intrónicas, proporcionando una herramienta eficaz para investigar los efectos de dichas variantes sobre la expresión y función de *GATA2*.
- 3.- La variante c.1017+572C>T produce una pérdida de expresión de *GATA2*, sin afectar al proceso de *splicing* del pre-ARNm de este gen.
- 4.- Los SNVs que afectan a los sitios canónicos de *splicing* poseen efectos deletéreos sobre regiones cruciales para la función del factor de transcripción *GATA2*.
- 5.- La variante c.1018-2A>C induce la pérdida de expresión de *GATA2*, así como la producción de 3 isoformas proteicas diferentes, dos de las cuales son truncadas. Esta variante afecta al exón 5, región que codifica para el dominio ZF2 de la proteína *GATA2*.
- 6.- La variante c.1143+2T>A causa la pérdida de expresión de *GATA2* e induce un patrón de *splicing* aberrante que produce el salto completo del exón 5, lo que se traduce en la pérdida completa del dominio ZF2 de la proteína *GATA2* y en la activación de un nuevo sitio donador a más de 100 pb, como se confirmó a nivel proteico con la producción de una isoforma truncada.
- 7.- Los ensayos *in vitro* son herramientas clave para evaluar el impacto funcional de las variantes genéticas.
- 8.- El uso de predictores *in silico* amplía la capacidad de los estudios funcionales para analizar el efecto de las alteraciones genéticas.

REFERENCIAS BIBLIOGRÁFICAS

1. Aktar, A., & Heit, B. (2023). Role of the pioneer transcription factor GATA2 in health and disease. *Journal of Molecular Medicine*, 101(10), 1191–1208. <https://doi.org/10.1007/s00109-023-02359-8>
2. Al Seraihi, A. F., Rio-Machin, A., Tawana, K., Bödör, C., Wang, J., Nagano, A., Heward, J. A., Iqbal, S., Best, S., Lea, N., McLornan, D., Kozyra, E. J., Wlodarski, M. W., Niemeyer, C. M., Scott, H., Hahn, C., Ellison, A., Tummala, H., Cardoso, S. R., ... Fitzgibbon, J. (2018). GATA2 monoallelic expression underlies reduced penetrance in inherited GATA2-mutated MDS/AML. *Leukemia*, 32(11), 2502–2507. <https://doi.org/10.1038/s41375-018-0134-9>
3. Arber, D. A., Orazi, A., Hasserjian, R. P., Borowitz, M. J., Calvo, K. R., Kvasnicka, H.-M., Wang, S. A., Bagg, A., Barbui, T., Branford, S., Bueso-Ramos, C. E., Cortes, J. E., Dal Cin, P., DiNardo, C. D., Dombret, H., Duncavage, E. J., Ebert, B. L., Estey, E. H., Facchetti, F., ... Tefferi, A. (2022). International Consensus Classification of Myeloid Neoplasms and Acute Leukemias: integrating morphologic, clinical, and genomic data. *Blood*, 140(11), 1200–1228. <https://doi.org/10.1182/blood.2022015850>
4. Arber, D. A., Orazi, A., Hasserjian, R., Thiele, J., Borowitz, M. J., Le Beau, M. M., Bloomfield, C. D., Cazzola, M., & Vardiman, J. W. (2016). The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*, 127(20), 2391–2405. <https://doi.org/10.1182/blood-2016-03-643544>
5. Balloux, F., Brønstad Brynildsrud, O., van Dorp, L., Shaw, L. P., Chen, H., Harris, K. A., Wang, H., & Eldholm, V. (2018). From Theory to Practice: Translating Whole-Genome Sequencing (WGS) into the Clinic. *Trends in Microbiology*, 26(12), 1035–1048. <https://doi.org/10.1016/j.tim.2018.08.004>
6. Bonnal, S. C., López-Oreja, I., & Valcárcel, J. (2020). Roles and mechanisms of alternative splicing in cancer — implications for care. *Nature Reviews Clinical Oncology*, 17(8), 457–474. <https://doi.org/10.1038/s41571-020-0350-x>
7. Calvo, K. R., & Hickstein, D. D. (2023). The spectrum of GATA2 deficiency syndrome. *Blood*, 141(13), 1524–1532. <https://doi.org/10.1182/blood.2022017764>
8. Churpek, J. E., & Bresnick, E. H. (2019). Transcription factor mutations as a cause of familial myeloid neoplasms. *Journal of Clinical Investigation*, 129(2), 476–488. <https://doi.org/10.1172/JCI120854>
9. Collin, M., Dickinson, R., & Bigley, V. (2015). Haematopoietic and immune defects associated with GATA2 mutation. *British Journal of Haematology*, 169(2), 173–187. <https://doi.org/10.1111/bjh.13317>
10. da Silva Lima, F., da Silva Gonçalves, C. E., & Fock, R. A. (2023). A review of the role of zinc finger proteins on hematopoiesis. *Journal of Trace Elements in Medicine and Biology*, 80, 127290. <https://doi.org/10.1016/j.jtemb.2023.127290>
11. Desviat, L. R., Pérez, B., & Ugarte, M. (2012). *Minigenes to Confirm Exon Skipping Mutations* (pp. 37–47). https://doi.org/10.1007/978-1-61779-767-5_3

12. Dickinson, R. E., Milne, P., Jardine, L., Zandi, S., Swierczek, S. I., McGovern, N., Cookson, S., Ferozepurwalla, Z., Langridge, A., Pagan, S., Gennery, A., Heiskanen-Kosma, T., Hämäläinen, S., Seppänen, M., Helbert, M., Tholouli, E., Gambineri, E., Reykdal, S., Gottfreðsson, M., ... Collin, M. (2014). The evolution of cellular deficiency in GATA2 mutation. *Blood*, *123*(6), 863–874. <https://doi.org/10.1182/blood-2013-07-517151>
13. Dobrzycki, T., Mahony, C. B., Krecsmarik, M., Koyunlar, C., Rispoli, R., Peulen-Zink, J., Gussinklo, K., Fedlaoui, B., de Pater, E., Patient, R., & Monteiro, R. (2020). Deletion of a conserved Gata2 enhancer impairs haemogenic endothelium programming and adult Zebrafish haematopoiesis. *Communications Biology*, *3*(1), 71. <https://doi.org/10.1038/s42003-020-0798-3>
14. Fabozzi, F., Strocchio, L., Mastronuzzi, A., & Merli, P. (2021). GATA2 and marrow failure. *Best Practice & Research Clinical Haematology*, *34*(2), 101278. <https://doi.org/10.1016/j.beha.2021.101278>
15. Feinberg, A. P., & Levchenko, A. (2023). Epigenetics as a mediator of plasticity in cancer. *Science*, *379*(6632). <https://doi.org/10.1126/science.aaw3835>
16. Ferrell, J. E. (2012). Bistability, Bifurcations, and Waddington's Epigenetic Landscape. *Current Biology*, *22*(11), R458–R466. <https://doi.org/10.1016/j.cub.2012.03.045>
17. Friedman, M. J., Wagner, T., Lee, H., Rosenfeld, M. G., & Oh, S. (2024). Enhancer–promoter specificity in gene transcription: molecular mechanisms and disease associations. *Experimental & Molecular Medicine*, *56*(4), 772–787. <https://doi.org/10.1038/s12276-024-01233-y>
18. Fujiwara, T. (2017). GATA Transcription Factors: Basic Principles and Related Human Disorders. *The Tohoku Journal of Experimental Medicine*, *242*(2), 83–91. <https://doi.org/10.1620/tjem.242.83>
19. Gaildrat, P., Killian, A., Martins, A., Tournier, I., Frébourg, T., & Tosi, M. (2010). *Use of Splicing Reporter Minigene Assay to Evaluate the Effect on Splicing of Unclassified Genetic Variants* (pp. 249–257). https://doi.org/10.1007/978-1-60761-759-4_15
20. Gu, M., Kovilakam, S. C., Dunn, W. G., Marando, L., Barcena, C., Mohorianu, I., Smith, A., Kar, S. P., Fabre, M. A., Gerstung, M., Cargo, C. A., Malcovati, L., Quiros, P. M., & Vassiliou, G. S. (2023). Multiparameter prediction of myeloid neoplasia risk. *Nature Genetics*, *55*(9), 1523–1530. <https://doi.org/10.1038/s41588-023-01472-1>
21. Haas, S., Trumpp, A., & Milsom, M. D. (2018). Causes and Consequences of Hematopoietic Stem Cell Heterogeneity. *Cell Stem Cell*, *22*(5), 627–638. <https://doi.org/10.1016/j.stem.2018.04.003>
22. Hanahan, D., & Weinberg, R. A. (2000). The Hallmarks of Cancer. *Cell*, *100*(1), 57–70. [https://doi.org/10.1016/S0092-8674\(00\)81683-9](https://doi.org/10.1016/S0092-8674(00)81683-9)
23. Harrison, S. M., Biesecker, L. G., & Rehm, H. L. (2019). Overview of Specifications to the ACMG/AMP Variant Interpretation Guidelines. *Current Protocols in Human Genetics*, *103*(1). <https://doi.org/10.1002/cphg.93>
24. Hewitt, K. J., Johnson, K. D., Gao, X., Keles, S., & Bresnick, E. H. (2016). *The Hematopoietic Stem and Progenitor Cell Cistrome* (pp. 45–76). <https://doi.org/10.1016/bs.ctdb.2016.01.002>

25. Hirabayashi, S., Wlodarski, M. W., Kozyra, E., & Niemeyer, C. M. (2017). Heterogeneity of GATA2-related myeloid neoplasms. *International Journal of Hematology*, *106*(2), 175–182. <https://doi.org/10.1007/s12185-017-2285-2>
26. Homan, C. C., Venugopal, P., Arts, P., Shahrin, N. H., Feurstein, S., Rawlings, L., Lawrence, D. M., Andrews, J., King-Smith, S. L., Harvey, N. L., Brown, A. L., Scott, H. S., & Hahn, C. N. (2021). GATA2 deficiency syndrome: A decade of discovery. *Human Mutation*, *42*(11), 1399–1421. <https://doi.org/10.1002/humu.24271>
27. Hryhorowicz, M., Lipiński, D., Zeyland, J., & Słomski, R. (2017). CRISPR/Cas9 Immune System as a Tool for Genome Engineering. *Archivum Immunologiae et Therapiae Experimentalis*, *65*(3), 233–240. <https://doi.org/10.1007/s00005-016-0427-5>
28. Hsu, A. P., McReynolds, L. J., & Holland, S. M. (2015). GATA2 deficiency. *Current Opinion in Allergy & Clinical Immunology*, *15*(1), 104–109. <https://doi.org/10.1097/ACI.0000000000000126>
29. Jung, M., Cordes, S., Zou, J., Yu, S. J., Guitart, X., Hong, S. G., Dang, V., Kang, E., Donaires, F. S., Hassan, S. A., Albitar, M., Hsu, A. P., Holland, S. M., Hickstein, D. D., Townsley, D., Dunbar, C. E., & Winkler, T. (2018). GATA2 deficiency and human hematopoietic development modeled using induced pluripotent stem cells. *Blood Advances*, *2*(23), 3553–3565. <https://doi.org/10.1182/bloodadvances.2018017137>
30. Katsanis, S. H., & Katsanis, N. (2013). Molecular genetic testing and the future of clinical genomics. *Nature Reviews Genetics*, *14*(6), 415–426. <https://doi.org/10.1038/nrg3493>
31. Khandekar, M., Brandt, W., Zhou, Y., Dagenais, S., Glover, T. W., Suzuki, N., Shimizu, R., Yamamoto, M., Lim, K.-C., & Engel, J. D. (2007). A Gata2 intronic enhancer confers its pan-endothelia-specific regulation. *Development*, *134*(9), 1703–1712. <https://doi.org/10.1242/dev.001297>
32. Khoury, J. D., Solary, E., Abla, O., Akkari, Y., Alaggio, R., Apperley, J. F., Bejar, R., Berti, E., Busque, L., Chan, J. K. C., Chen, W., Chen, X., Chng, W.-J., Choi, J. K., Colmenero, I., Coupland, S. E., Cross, N. C. P., De Jong, D., Elghetany, M. T., ... Hochhaus, A. (2022). The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms. *Leukemia*, *36*(7), 1703–1719. <https://doi.org/10.1038/s41375-022-01613-1>
33. Kotmayer, L., Romero-Moya, D., Marin-Bejar, O., Kozyra, E., Català, A., Bigas, A., Wlodarski, M. W., Bödör, C., & Giorgetti, A. (2022). GATA2 deficiency and MDS/AML: Experimental strategies for disease modelling and future therapeutic prospects. *British Journal of Haematology*, *199*(4), 482–495. <https://doi.org/10.1111/bjh.18330>
34. Kozyra, E. J., Pastor, V. B., Lefkopoulos, S., Sahoo, S. S., Busch, H., Voss, R. K., Erlacher, M., Lebrecht, D., Szvetnik, E. A., Hirabayashi, S., Pasaulienė, R., Pedace, L., Tartaglia, M., Klemann, C., Metzger, P., Boerries, M., Català, A., Hasle, H., de Haas, V., ... Wlodarski, M. W. (2020). Synonymous GATA2 mutations result in selective loss of mutated RNA and are common in patients with GATA2 deficiency. *Leukemia*, *34*(10), 2673–2687. <https://doi.org/10.1038/s41375-020-0899-5>

35. Leubolt, G., Redondo Monte, E., & Greif, P. A. (2020). GATA2 mutations in myeloid malignancies: Two zinc fingers in many pies. *IUBMB Life*, 72(1), 151–158. <https://doi.org/10.1002/iub.2204>
36. Liu, H., & Naismith, J. H. (2008). An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnology*, 8(1), 91. <https://doi.org/10.1186/1472-6750-8-91>
37. McClure, R. F., Ewalt, M. D., Crow, J., Temple-Smolkin, R. L., Pullambhatla, M., Sargent, R., & Kim, A. S. (2018). Clinical Significance of DNA Variants in Chronic Myeloid Neoplasms. *The Journal of Molecular Diagnostics*, 20(6), 717–737. <https://doi.org/10.1016/j.jmoldx.2018.07.002>
38. Mehta, C., Johnson, K. D., Gao, X., Ong, I. M., Katsumura, K. R., McIver, S. C., Ranheim, E. A., & Bresnick, E. H. (2017). Integrating Enhancer Mechanisms to Establish a Hierarchical Blood Development Program. *Cell Reports*, 20(12), 2966–2979. <https://doi.org/10.1016/j.celrep.2017.08.090>
39. Mir, M. A., Kochuparambil, S. T., Abraham, R. S., Rodriguez, V., Howard, M., Hsu, A. P., Jackson, A. E., Holland, S. M., & Patnaik, M. M. (2015). Spectrum of myeloid neoplasms and immune deficiency associated with germline GATA2 mutations. *Cancer Medicine*, 4(4), 490–499. <https://doi.org/10.1002/cam4.384>
40. Naberezhnov, D. S., Lesovaya, E. A., Kirsanov, K. I., & Yakubovskaya, M. G. (2021). *Artificial introns for effective expression of transgenes in mammalian cells (Preprint)*. <https://doi.org/10.1101/2021.09.01.457939>
41. Oleaga-Quintas, C., de Oliveira-Júnior, E. B., Rosain, J., Rapaport, F., Deswarte, C., Guérin, A., Sajjath, S. M., Zhou, Y. J., Marot, S., Lozano, C., Branco, L., Fernández-Hidalgo, N., Lew, D. B., Brunel, A.-S., Thomas, C., Launay, E., Arias, A. A., Cuffel, A., Monjo, V. C., ... Bustamante, J. (2021). Inherited GATA2 Deficiency Is Dominant by Haploinsufficiency and Displays Incomplete Clinical Penetrance. *Journal of Clinical Immunology*, 41(3), 639–657. <https://doi.org/10.1007/s10875-020-00930-3>
42. Peters, I. J. A., de Pater, E., & Zhang, W. (2023). The role of GATA2 in adult hematopoiesis and cell fate determination. *Frontiers in Cell and Developmental Biology*, 11. <https://doi.org/10.3389/fcell.2023.1250827>
43. Putscher, E., Hecker, M., Fitzner, B., Lorenz, P., & Zettl, U. K. (2021). Principles and Practical Considerations for the Analysis of Disease-Associated Alternative Splicing Events Using the Gateway Cloning-Based Minigene Vectors pDESTsplice and pSpliceExpress. *International Journal of Molecular Sciences*, 22(10), 5154. <https://doi.org/10.3390/ijms22105154>
44. Rafei, H., & DiNardo, C. D. (2019). Hereditary myeloid malignancies. *Best Practice & Research Clinical Haematology*, 32(2), 163–176. <https://doi.org/10.1016/j.beha.2019.05.001>
45. Rajput, R. V., & Arnold, D. E. (2023). GATA2 Deficiency: Predisposition to Myeloid Malignancy and Hematopoietic Cell Transplantation. *Current Hematologic Malignancy Reports*, 18(4), 89–97. <https://doi.org/10.1007/s11899-023-00695-7>
46. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., & Rehm, H. L. (2015). Standards and guidelines for the

- interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, 17(5), 405–424. <https://doi.org/10.1038/gim.2015.30>
47. Sahoo, S. S., Kozyra, E. J., & Wlodarski, M. W. (2020). Germline predisposition in myeloid neoplasms: Unique genetic and clinical features of GATA2 deficiency and SAMD9/SAMD9L syndromes. *Best Practice & Research Clinical Haematology*, 33(3), 101197. <https://doi.org/10.1016/j.beha.2020.101197>
 48. Santiago, M., Liquori, A., Such, E., Zúñiga, Á., & Cervera, J. (2023). The Clinical Spectrum, Diagnosis, and Management of GATA2 Deficiency. *Cancers*, 15(5), 1590. <https://doi.org/10.3390/cancers15051590>
 49. Shamriz, O., Zahalka, N., Simon, A. J., Lev, A., Barel, O., Mor, N., Tal, Y., Segel, M. J., Somech, R., Yonath, H., & Toker, O. (2022). GATA2 Deficiency in Adult Life Is Characterized by Phenotypic Diversity and Delayed Diagnosis. *Frontiers in Immunology*, 13. <https://doi.org/10.3389/fimmu.2022.886117>
 50. Shimizu, R., & Yamamoto, M. (2020). Quantitative and qualitative impairments in GATA2 and myeloid neoplasms. *IUBMB Life*, 72(1), 142–150. <https://doi.org/10.1002/iub.2188>
 51. Soukup, A. A., & Bresnick, E. H. (2020). GATA2 +9.5 enhancer: from principles of hematopoiesis to genetic diagnosis in precision medicine. *Current Opinion in Hematology*, 27(3), 163–171. <https://doi.org/10.1097/MOH.0000000000000576>
 52. Suzuki, M., Katayama, S., & Yamamoto, M. (2020). Two effects of GATA2 enhancer repositioning by 3q chromosomal rearrangements. *IUBMB Life*, 72(1), 159–169. <https://doi.org/10.1002/iub.2191>
 53. Tremblay, M., Sanchez-Ferras, O., & Bouchard, M. (2018). GATA transcription factors in development and disease. *Development*, 145(20). <https://doi.org/10.1242/dev.164384>
 54. Tronick, E., & Hunter, R. G. (2016). Waddington, Dynamic Systems, and Epigenetics. *Frontiers in Behavioral Neuroscience*, 10. <https://doi.org/10.3389/fnbeh.2016.00107>
 55. Walker, L. C., Hoya, M. de la, Wiggins, G. A. R., Lindy, A., Vincent, L. M., Parsons, M. T., Canson, D. M., Bis-Brewer, D., Cass, A., Tchourbanov, A., Zimmermann, H., Byrne, A. B., Pesaran, T., Karam, R., Harrison, S. M., Spurdle, A. B., Biesecker, L. G., Harrison, S. M., Tayoun, A. A., ... Topper, S. (2023). Using the ACMG/AMP framework to capture evidence related to predicted and observed impact on splicing: Recommendations from the ClinGen SVI Splicing Subgroup. *The American Journal of Human Genetics*, 110(7), 1046–1067. <https://doi.org/10.1016/j.ajhg.2023.06.002>
 56. Wang, H., Zhou, N., Ding, F., Li, Z., Chen, R., Han, A., & Liu, R. (2011). An efficient approach for site-directed mutagenesis using central overlapping primers. *Analytical Biochemistry*, 418(2), 304–306. <https://doi.org/10.1016/j.ab.2011.07.008>
 57. Wang, X., Muramatsu, H., Okuno, Y., Sakaguchi, H., Yoshida, K., Kawashima, N., Xu, Y., Shiraishi, Y., Chiba, K., Tanaka, H., Saito, S., Nakazawa, Y., Masunari, T., Hirose, T., Elmahdi, S., Narita, A., Doisaki, S., Ismael, O., Makishima, H., ... Kojima, S. (2015). GATA2 and secondary mutations in familial myelodysplastic syndromes and pediatric myeloid malignancies. *Haematologica*, 100(10), e398–e401. <https://doi.org/10.3324/haematol.2015.127092>

58. You, X., Chang, Y.-I., Kong, G., Ranheim, E. A., Zhou, Y., Johnson, K. D., Mehta, C., Bresnick, E. H., & Zhang, J. (2021). Gata2 -77 enhancer regulates adult hematopoietic stem cell survival. *Leukemia*, 35(3), 901–905. <https://doi.org/10.1038/s41375-020-0942-6>
59. You, X., Zhou, Y., Chang, Y.-I., Kong, G., Ranheim, E. A., Johnson, K. D., Soukup, A. A., Bresnick, E. H., & Zhang, J. (2022). Gata2 +9.5 enhancer regulates adult hematopoietic stem cell self-renewal and T-cell development. *Blood Advances*, 6(4), 1095–1099. <https://doi.org/10.1182/bloodadvances.2021004311>
60. Yu, V. W. C., & Scadden, D. T. (2016). *Hematopoietic Stem Cell and Its Bone Marrow Niche* (pp. 21–44). <https://doi.org/10.1016/bs.ctdb.2016.01.009>
61. Zhang, K., Yin, X., Shi, K., Zhang, S., Wang, J., Zhao, S., Deng, H., Zhang, C., Wu, Z., Li, Y., Zhou, X., & Deng, W. (2021). A high-efficiency method for site-directed mutagenesis of large plasmids based on large DNA fragment amplification and recombinational ligation. *Scientific Reports*, 11(1), 10454. <https://doi.org/10.1038/s41598-021-89884-z>
62. Zheng, L. (2004). An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Research*, 32(14), e115–e115. <https://doi.org/10.1093/nar/gnh110>

ANEXOS, FIGURAS Y TABLAS SUPLEMENTARIAS

Tabla Suplementaria 1. Listado de variantes sobre *GATA2* reportadas en pacientes con deficiencia en *GATA2*.

ENST00000341105.2:c.1054T>C	ENST00000341105.2:c.1084C>T	ENST00000341105.2:c.312_313dup
ENST00000341105.2:c.1054T>G	ENST00000341105.2:c.1085G>C	ENST00000341105.2:c.317_318del
ENST00000341105.2:c.1055G>T	ENST00000341105.2:c.1099del	ENST00000341105.2:c.318dup
ENST00000341105.2:c.1009C>T	ENST00000341105.2:c.1099dup	ENST00000341105.2:c.351C>G
ENST00000341105.2:c.1017+1del	ENST00000341105.2:c.1103_1104del	ENST00000341105.2:c.353del
ENST00000341105.2:c.1017+2T>C	ENST00000341105.2:c.1110C>G	ENST00000341105.2:c.404dup
ENST00000341105.2:c.1017+2T>G	ENST00000341105.2:c.1113C>A	ENST00000341105.2:c.414_417del
ENST00000341105.2:c.1017+513_1017+540del	ENST00000341105.2:c.1113C>G	ENST00000341105.2:c.416_417del
ENST00000341105.2:c.1017+572C>T	ENST00000341105.2:c.1113del	ENST00000341105.2:c.417dup
ENST00000341105.2:c.1018_1028del	ENST00000341105.2:c.1114G>A	ENST00000341105.2:c.423C>A
ENST00000341105.2:c.1018-10_1037del	ENST00000341105.2:c.1117_1131del	ENST00000341105.2:c.437del
ENST00000341105.2:c.1018-11_1027del	ENST00000341105.2:c.1117T>C	ENST00000341105.2:c.538G>T
ENST00000341105.2:c.1018-1G>A	ENST00000341105.2:c.1118G>A	ENST00000341105.2:c.561dup
ENST00000341105.2:c.1018-1G>T	ENST00000341105.2:c.1123C>T	ENST00000341105.2:c.579dup
ENST00000341105.2:c.1018-2A>G	ENST00000341105.2:c.1124del	ENST00000341105.2:c.586_587insTCTTCCGC
ENST00000341105.2:c.1018-2A>T	ENST00000341105.2:c.1126_1127insTACTACAA	ENST00000341105.2:c.58C>T
ENST00000341105.2:c.1018-3_1031del	ENST00000341105.2:c.1128C>G	ENST00000341105.2:c.593del
ENST00000341105.2:c.1021_1022insGCCGACAGAG	ENST00000341105.2:c.1129T>G	ENST00000341105.2:c.599del
ENST00000341105.2:c.1021_1024del	ENST00000341105.2:c.1132A>T	ENST00000341105.2:c.599dup
ENST00000341105.2:c.1021_1031del	ENST00000341105.2:c.1142del	ENST00000341105.2:c.610C>T
ENST00000341105.2:c.1021del	ENST00000341105.2:c.1143+2T>A	ENST00000341105.2:c.627_628insCGGC
ENST00000341105.2:c.1023_1024insCGCCAGAAGAGCCGGC	ENST00000341105.2:c.1143+5G>A	ENST00000341105.2:c.670G>T
ENST00000341105.2:c.1023del	ENST00000341105.2:c.1143+5G>C	ENST00000341105.2:c.685del
ENST00000341105.2:c.1024_1026del	ENST00000341105.2:c.1154C>A	ENST00000341105.2:c.710del
ENST00000341105.2:c.1025_1026insGCCG	ENST00000341105.2:c.1160C>A	ENST00000341105.2:c.735dup
ENST00000341105.2:c.1031_1049del	ENST00000341105.2:c.1162A>G	ENST00000341105.2:c.769_770insGCCACGACT
ENST00000341105.2:c.1035_1036insTCTGGCC	ENST00000341105.2:c.1163T>C	ENST00000341105.2:c.77A>C
ENST00000341105.2:c.1035_1036insCGGC	ENST00000341105.2:c.1168A>G	ENST00000341105.2:c.793_802del
ENST00000341105.2:c.1041del	ENST00000341105.2:c.1172_1175del	ENST00000341105.2:c.802G>T
ENST00000341105.2:c.1045T>C	ENST00000341105.2:c.1186C>T	ENST00000341105.2:c.817_818del
ENST00000341105.2:c.1045T>G	ENST00000341105.2:c.1187G>A	ENST00000341105.2:c.818dup
ENST00000341105.2:c.1046G>T	ENST00000341105.2:c.1187G>T	ENST00000341105.2:c.857C>T

ENST00000341105.2:c.1052A>G	ENST00000341105.2:c.1192C>T	ENST00000341105.2:c.869C>A
ENST00000341105.2:c.1054del	ENST00000341105.2:c.1193G>A	ENST00000341105.2:c.892dup
ENST00000341105.2:c.1060A>C	ENST00000341105.2:c.1200_1201insGTCCAACAAGTCCAAGA	ENST00000341105.2:c.915_916del
ENST00000341105.2:c.1061C>G	ENST00000341105.2:c.1281dup	ENST00000341105.2:c.917G>A
ENST00000341105.2:c.1061C>T	ENST00000341105.2:c.130G>T	ENST00000341105.2:c.937C>T
ENST00000341105.2:c.1065_1067del	ENST00000341105.2:c.1322_1323insGCCA	ENST00000341105.2:c.941_951del
ENST00000341105.2:c.1069A>G	ENST00000341105.2:c.1339A>C	ENST00000341105.2:c.941_942insACCTGTGCAAT
ENST00000341105.2:c.1070C>T	ENST00000341105.2:c.1341C>A	ENST00000341105.2:c.944T>C
ENST00000341105.2:c.1072_1074del	ENST00000341105.2:c.161C>A	ENST00000341105.2:c.952G>A
ENST00000341105.2:c.1073C>A	ENST00000341105.2:c.17_18del	ENST00000341105.2:c.956_962del
ENST00000341105.2:c.1075T>G	ENST00000341105.2:c.177C>G	ENST00000341105.2:c.968dup
ENST00000341105.2:c.1076T>C	ENST00000341105.2:c.206_208del	ENST00000341105.2:c.971_972insAGATGAATGGGCAGAACCGACCACA
ENST00000341105.2:c.1078T>A	ENST00000341105.2:c.207_208del	ENST00000341105.2:c.982C>T
ENST00000341105.2:c.1079G>T	ENST00000341105.2:c.232dup	ENST00000341105.2:c.988C>T
ENST00000341105.2:c.1081C>G	ENST00000341105.2:c.256del	ENST00000341105.2:c.989_990insGACC
ENST00000341105.2:c.1081C>T	ENST00000341105.2:c.257_258del	-
ENST00000341105.2:c.1082G>A	ENST00000341105.2:c.302del	-
ENST00000341105.2:c.1082G>T	ENST00000341105.2:c.303del	-
ENST00000341105.2:c.1084_1095del	ENST00000341105.2:c.306del	-

N=149 variantes. El listado incluye variantes clasificadas como patogénicas, potencialmente patogénicas y de significado incierto descritas en pacientes afectados por NM con implicación genética de GATA2. El conjunto de variantes se puede encontrar en el material suplementario de la revisión *GATA2 deficiency syndrome: A decade of discovery* (DOI: [10.1002/humu.24271](https://doi.org/10.1002/humu.24271)).

Anexo 1. Diseño químico de GATA2 para el ORF del vector de expresión

>3 dna:chromosome:GRCh38:3:128479377:128493251:-1 – MODIFIED FOR VECTOR with small introns AND without alternative exons.

Exon 1

GAGAGAGAGGCAGCGTGAGCGCCAGGAAGGTAGCGAGGCCAGCGTCGCCCGGGACTCGCTGCTCAAGTCTGTCTATTGCCTGCCGCCACATCCATCCTAGCAGGGCCCCGTCGCCACCAGGCGGACAAAAGCGGTCCGCTGAACACCATGCGGCCGCTCGGC GTGCCGCCAGGCTCTGCTGGTGTAGCGCCGCCACCCGCGCCAGGTCCCGCGAGCCCGCCTGCCGCGACCTCGCCCTGCTC CCAGCTACTCCAGGCCCGTCCGCCCGGGGCGCCGCCACCAGCGCCTCGCTCG

Intron 1 (chimeric)

Gtgagtttctccacttccacttcccttggcccgcccttccgccccggccccgcccgcctggcaccgcaatcgcttggctcgtttgcccctgtggccctaccttgg ggtcgccttggccctgccaagaccgaaaccctggttatagggacttaacagaggagatttggttacctgggcacagcaggccccctcggcctcttctt cccatttctcgagccacagccacagactccccggcgctgaccatgtgcacgggtgtgtattctggagccggtcaccacgtgagtgctgctgactgaacc ccctccccgccttcttctgtttgagccttggccttctccaccgggactggtgctcttctcgggatctgggctggggctccgtggcgtcgggacacctgtg gtgggacttgggggtgtcaggcgctggcgccacgctcactcccccttctcctgcgcag

Exon 2 + EGFP ADNC

GGCCGTTGCCGTCTGCACCCAGACCCTGAGCCGCCGCCGCCGCCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGGTG GTGCCCATCCTGGTCTGAGCTGGACGGCGACGTAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCT ACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGCAAGCTGCCCGTGCCTGGCCACCCTCGTGACCACCTGACCTAC GGCCTGCAGTGCTTACGCCGCTACCCCGACCACATGAAGCAGCAGACTTCTTCAAGTCCGCCATGCCGAAGGCTACGTCCA GGAGCGACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAAC CGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGACAAGCTGGAGTACAACCTACAACAGCCACA ACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAAGTTCAGATCCGCCACAACATCGAGGACGGCAGCGTG CAGCTCGCCGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCCGTGCTGCTGCCGACAACCACTACCTGAGCACCCA

GTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCG
GCATGGACGAGCTGTACAAGATGGAGGTGGCGCCGAGCA
GCCGCGCTGGATGGCGCACCCGGCCGTGCTGAATGCGCAGCACCCGACTCACACCACCCGGCCTGGCGCACAACTACATG
GAACCCGCGCAGCTGCTCCTCCAGACGAGGTGGACGTCTTCAATCACTCGACTCGCAGGGCAACCCCTACTATGCCAAC
CCCGCTCACGCGCGGGCGCGCTCTCTACAGCCCCGCGCACG

Intron 2 (chimeric)

Gtgagactgggcccgtggtgatgagaaccaggcgcgcgcccaggcggaggaggaggaggcccgtctgcttcccgatgtaggatccgcaggaat
cgagctgtgaaaaattggggcgggagaggtgggagcagcggccgattggggagggtctgggaccacagggttctgccactccagctggcctgtgagggttcc
ctgtagggctgtccggtggggctctctatgcccactgtcttccagctggactgacattcctgattattaccggtgggtattatgttctggctttttggggaggga
tgactgtggttctgggagctgtgatctcaatgtctgtcaggggcgtccctagctctgctacctgatcttctgccaccctgatcctctctcttggcccag

Exon 3

CCCGCCTGACCGGAGGCCAGATGTGCCGCCACACTTGTGACAGCCCGGTTTGCCTGGCTGGACGGGGGCAAA
GCAGCCCTCTGCGCTGCGGCCACCACCACAACCCCTGGACCGTGAGCCCTTCTCAAAGACGCCACTGCACCCCTCAGCT
GCTGGAGCCCTGGAGGCCACTCTGTGTACCCAGGGGCTGGGGTGGGAGCGGGGAGGCAGCGGGAGCTCAGTGGC
CTCCCTACCCCTACAGCAGCCACTCTGGCTCCACCTTTTCGGCTTCCACCCAGCCACCAAAGAAGTGTCTCTGACCT
AGCACACGCGGGCTGCGTCCAGCCTCATTTCCGCGGGGGTAGTGCAGCCGAGGAGAGGACAAGGACGCGCTCAAG
TACCAGGTGTCACTGACGAGGAGCATGAAGATGGAAGTGGCAGTCCCCTGCGCCAGGCCTAGCTACTAGGGCACCCAGCC
TGCTACACACCCCTACCCCTACCCCTCCTATGTGCCGCGGCTGCCACGACTACAGCAGCGGACTTCCACCCCGGA
GGCTTCTGGGGGACCGCCTCCAGCTTACCCCTAAGCAGCGCAGCAAGGCTCGTTCCTGTTACG

Intron 3 (chimeric)

gtaaaggcaggtgctgggactctgtggaagaggggagcatttgcgtttttgtggtggggagctgtgactgggagaggtggcagtggttccagatctggggat
gggtgatgtctcgtctaataagccccagcagatgtttgagacccccgggtcaggagagagagactctaaaactttgccatttttaagtctccagtgtaacccca
aagttcagttcctgccccagggttgctcctcttccatccctcctgtgaggattggaagcaggggattggcttaaacgacttggaacctgtccccctccatgt
gggagaccctctgctcctctctgcccaggctgttcagccaggcccaggccaacctgtgctgagaggaccggcactgctcggctggcttctctcctctgg
actcctcccagaaactgcccgttaagcaggccccctgtctctcctgttccctgcag

Exon 4

AAGGCCGGGAGTGTGTAACCTGTGGGGCCACAGCCACCCCTCTGTGGCGCGGGACGGCACCGGCCACTACCTGTGCAATGC
CTGTGGCCTTACCACAAGATGAATGGGCAGAACCGACCACTCATCAAGCCCAAGCGAAGACTG

Intron 4 (full intronic sequence)

gtaggaggggacaggtggctgggaggggctgctggcaggagctggcggttaattacaggggaaaaaactcctcaaatgcagacgcttccgctgaaat
cctctttatcatgaaagcactgggatgtcagttgggtcgtctctttctggccagattcttgggagcagattctcctcggtataggagcccaccgggacacct
tgcgccacccactctgctccgggatccccgaagttgagtgccacgggcccagactcctgtccttggcctctgcttagctcttttaaaataggccatgaagtactt
ttccttggctcagccctcccagacccccgctcacaagctcctgcttatttaataaaacacaaactcaacccggccactaaaaaacctgccccttattatttt
tccatggagtcacctatactgtatatttcaattgagtgatttaaaaaaatgccttccgactcctgcccggagtttctatccggacatctgcagcctgaagataagga
aactcgtgatctgtttccgactctgcgagttttagagtcctcagctcagctcctgctcctgctgggctgtttgaaatttcaataacctcactctgcaataatgc
gtaaaatgtaagaataataatataatattttagggcgaagtgattatgaggttaaatcttccctgcttggggcctttttccctggagcagggcggggtga
ggccccgggtgggtagaggtggaggacgcggcgttgcccctgagtcagaattccagctcaggtgcttactaccctccctgccccgcggtgagctccctct
gtccctctgtgaccaggcttggcctgggctgttccaggctctgcagccctcagccccagccccacactcaccactggtgactcccctcagttctctggg
aagtttgggggacccccctgtcactgtgggctggcgttggggaaccggg
agagggatctgttttctgggtaaagcctcctcagcttctctgcaaggaccagcgctcattccagaccctacctctccaggcatttctgaggactaggact
cagaggggctcggggtgggtaagcttaaggttggggtatgggggctggatggggggatcagactcacatcagctggagagatggaagttctgtgtct
gcaactcccactgtgtagccctggccacatgtaattgatcactgaaatgtgctcgtgcaattgagggactgggttttaatttgaattgttagtagatctt
atltaaatggctcctgtgcccagctgtacagttggcaggtgcagctcgcactgtaaacctgcgctggcctcagcagactgactaccaggattatggatt
tgagcggagctgtgtagaggagacagaaatcgccccagatccaggggctcaggggggaccaagccgctcagcctcaggtatcctgtgtagagagccct
tctcagggcctcagtttcccattatggagttagagcgcagggtagttggggaggttagctaattctcctgtagctctgcaatccgttgattcaacatcaggctc
tgagagttcttattcaaaagttctgtgagcttctgacttattctcctcaaatctaaattcctggtctgagatgcttggattccatgtgagattagccctctgactg
agctggtgggactgggggtggagcgaagggtcagggaggggggtcaggtgggctgggagtcagcctctgacgctgcttgcctcccag

Exon 5

TCGGCCGCAAGAGCCGGCACCTGTTGTGCAATTGTGACAGCACAACCACACCTTATGGCGCGAAACGCCAACGGGG
ACCCTGTCTGCAACGCTGTGGCCTTACTACAAGCTGCACAAT

Intron 5 (full intronic sequence)

Gtgagtgcgccccccccggccacccccccccctccaggggacctctgcctttgtgctccaggcaagagggccccagccacaatatccagcttggcttggctggg
aagctgctccctgagtgagccagaagggtctccgtaagaggggtccttgcctctgctcaggaggtggagctggctaggacagggctctggactagggaagt
ggtttctctgcttaaaagggtcaggggtggggggaggactcagttggctggcagctgctggcagcgggtggcagagccagggaggggtgtgggtcagccccat
gccagaaccgccccctggaatgtagcatctggtgatgggactatgaaggtgggcacaaattcctggcttctgggacccctcagcttaccctgctctgtccac
gctgtggcggggtgggaggaatgttctgagggaaggaaactggccctctgaaaactggtggttgcctctag

Exon 6

GTAAACAGGCCACTGACCATGAAGAAGGAAGGGATCCAGACTCGGAACCGGAAGATGTCCAACAAGTCCAAGAAGAGCAAG
AAAGGGGCGGAGTGCTTCGAGGAGCTGTCAAAGTGCATGCAGGAGAAGTCATCCCCCTCAGTGCAGCTGCCCTGGCTGGAC
ACATGGCACCTGTGGCCACCTCCGCCCCCTCAGCCACTCCGGACACATCCTGCCACTCCGACGCCATCCACCCCTCCTCCAG

CCTCTCTTCG GCCACCCACCCGTC CAGCATGGTG ACCGCCATGGGCTAGGGAACAGATGGACGTCGAGGACCGGGCACTC
 CCGGGATGGGTGGACAAACCTTAGCAGCCAGCATTTCCCGAAGGCCGACACCCTCTGCCAGCCCGCTCGGGCCAGC
 ACCCCCTCTCTGGAGGGCGCCAGCAGCCTGCCAGCAGTTACTGTGAATGTTCCACCGCTGAGAGGCTGCCTCCGCACCT
 GACCGCTGCCAGGTGGGGTTTCTGCATGGACAGTTGTTGGAGAACAACAAGGACAACCTTTATGTAGAGAAAAGGAGGGG
 ACGGGACAGACGAAGGCAACCATTTTTAGAAGGAAAAGGATTAGGCCAAAATAATTTATTTTGTCTTGTCTTCTAACAAGGAC
 TTGGAGACTTGGTGGTCTGAGCTGTCCCAAGTCTCCGGTCTTCTCTCGGGATTGGCGGGTCCACTTGCCAGGGCTCTGGGGG
 CAGATTTGTGGGGACCTCAGCCTGCACCCTTCTCTCTGGCTTCCCTCTCTGAAATAGCCGAACTCCAGGCTGGGCTGAGCC
 AAAGCCAGAGTGGCCACGGCCAGGGAGGGTGAGCTGGTGCCTGCTTTGACGGGCCAGGCCCTGGAGGGCAGAGACAATC
 ACGGGCGGTCTGCACAGATTCCAGGCCAGGGCTGGGTACAGGAAGGAAACAACATTTTCTTAAAGGGGAAACGTCTCC
 CAGATCGCTCCCTTGGCTTGGAGCCGAAGCTGCTGTGACTGTGTCCCTTACTGAGCGCAAGCCACAGCCTGCTTGTGAGGT
 GGACCCTGTAAATACATCTTTTTCTGTAAACCTTCAACCCCTCGCCTCTACTCTGAGACAAAAGAAAAATATTAATAAAT
 GCATAGGCTTAACGCTGATGAGTTAATGTTTTATTTTAAACTCTTTTTGGGTCCAGTTGATTGTACGTAGCCACAGGAGCCC
 TGCTATGAAAGGAATAAACCTACACACAAGGTTGGAGCTTGAATCTTTTTGGAAAAGAGCTGGGATCCACAGCCCTAGT
 ATGAAAGCTGGGGTGGGGAGGGGCTTTGCTGCCCTGGTTCTGGGGCTGGTTGGCATTGTCTGGCCTGGCAGGGGGT
 GAAGGCAGGAGTTGGGGCAGGTACAGACCAGGACCCAGGAGAGGCTGTGTCCCTGCTGGGGTCTCAGTCCAGCTTTAC
 TGTGGCTGTCTGGATCCTTCCAAGGTACAGCTGTATATAACGTGTCCCGAGCTTAGATTCTGTATCGGTGACGGCGGGGTGT
 GGTGGCCTGTGAGGGCCCCAGGCCAGGAGGAGGATTGTGCTGATGTAGTGACCAATGCAATATGGGGCGGGCAGTCGCTG
 CAGGGAGACCACGGCCAGAAGTAACCTATTTTGTACTAGTGTCCGCATAAGAAAAAGAATCGGCAGTATTTTCTGTTTTATGT
 TTTATTTGGCTGTTTTATTTTGGATTAGTAACTAAGTTATTGTTAATTATGTACAACATTTATATATTGTCTGAAAAAATGTATGC
 TATCTCTTATTCCTTTAAAGTGAGTACTGTTAAGAATAATAAATACTTTTTGTGAA

Tabla Suplementaria 2. Parámetros fisicoquímicos de los cebadores diseñados para la amplificación de *GATA2*

		Longitud (pb)	GC (%)	T _m calc. (°C)
Para la validación a nivel de secuencia				
<i>Int1_F</i>	CATGGTGAGCAAGGGCGAG	19	63	62
<i>Int1_R</i>	CTCGCCCTTGCTCACCATG	19	63	62
<i>Int2_F</i>	CGTCTCCAGCCTCATCTTC	19	58	60
<i>Int2_R</i>	GAAGATGAGGCTGGAGACG	19	58	60
<i>Int3_F</i>	GAGTCAGAATTCCAGCTTCAG	21	48	62
<i>Int3_R</i>	CTGAAGCTGGAATTCTGACTC	21	48	62
<i>Int4_F</i>	CTCCTTGACTGAGCTGGTG	19	58	60
<i>Int4_R</i>	CACCAGCTCAGTCAAGGAG	19	58	60
<i>EGFP_F</i>	ACAACACTACAACAGCCACAACG	21	48	62
<i>E24_F</i>	CTGGCGCACAACACTACATGG	19	58	60
<i>E24_R</i>	GATGAGTGGTCCGGTTCTGC	19	58	60
Para la validación específica de las variantes				
<i>I45_F</i>	GGAGCTGGCGGTTAATTACAG	21	52	64
<i>I45_R</i>	TCTGAGTCCTAGTCCCTCAG	20	55	62
<i>E5_F</i>	CAGCTGCTACAGTGTGGAC	20	55	62
<i>E5_R</i>	GTCAGTGGCCTGTTAACCTA	20	50	60
Para el estudio de proceso de splicing alrededor del exón 5				
<i>E46_F</i>	CGGGAGTGTGTCAACTGTG	19	58	60
<i>E46_R</i>	CACCATGCTGGACGGGTG	18	67	60

El cálculo de la temperatura de fusión (T_m) se realizó con el modelo $T_m = 4 (G+C) + 2 (A+T)$. El valor real de la T_m depende de múltiples factores, como la tendencia a formar estructuras secundarias o la concentración de sales en el seno de la disolución del cebador, entre otros. Se diseñaron oligonucleótidos de, al menos, 60 °C de T_m para evitar cebadores con hibridaciones inespecíficas.

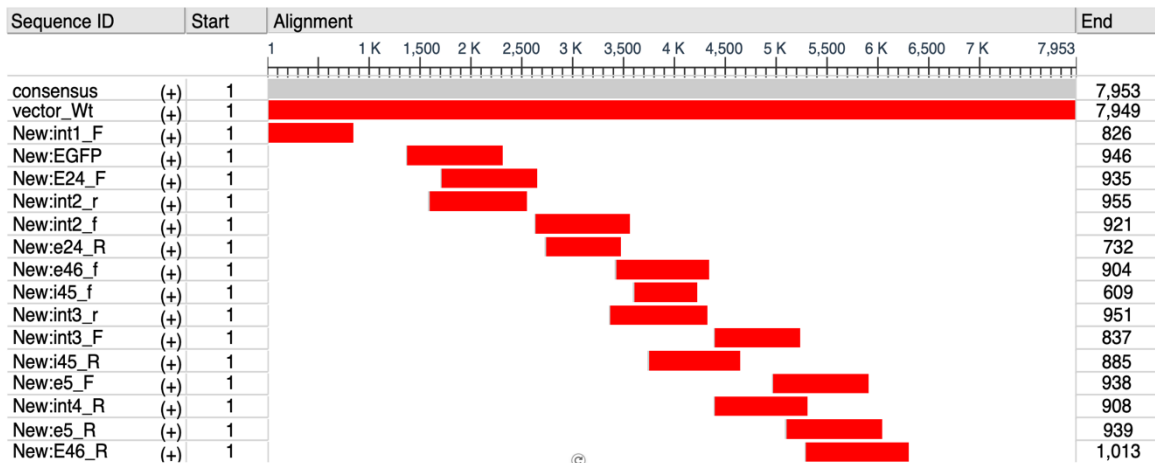


Figura Suplementaria 1. Alineamiento de las lecturas obtenidas mediante Sanger frente a la secuencia del ORF del vector. El alineamiento se realizó con el software online MAFFT, se descargó el archivo en formato .fasta y se visualizó con el NCBI MSA Viewer (versión 1.25.0). Asimismo, la secuencia se visualizó a nivel nucleotídico con el programa MEGA.

Tabla Suplementaria 4. Cebadores autosolapantes con los cambios nucleotídicos de interés para la mutagénesis

Nombre del cebador	Secuencia nucleotídica
GATA2_c.1143+5G>C_F	caagctgcacaatgtgactgcgccccg
GATA2_c.1143+5G>C_R	cggggcgagtcacattgtgcagcttg
GATA2_c.1143+2T>A_F	aagctgcacaatgagagtgcgccccgc
GATA2_c.1143+2T>A_R	gcgggggcgactctcattgtgcagctt
GATA2_c.1018-1G>A_F	tgcttgcctccaatcggccgc
GATA2_c.1018-1G>A_R	gcgccgattgggagggcaaggca
GATA2_c.1018-1G>T_F	tgcttgcctcccattcggccgc
GATA2_c.1018-1G>T_R	gcgccgaatgggagggcaaggca
GATA2_c.1018-2A>T_F	cttgcctccctgtcgccgcca
GATA2_c.1018-2A>T_R	tggcggccgacagggagggcaag
GATA2_c.1018-2A>G_F	cttgcctcccgtcgccgcca
GATA2_c.1018-2A>G_R	tggcggccgaccgggagggcaag
GATA2_c.1018-2A>C_F	cttgcctcccgtcgccgcca
GATA2_c.1018-2A>C_R	tggcggccgacgggagggcaag
GATA2_c.1017+572C>T_F	aacttcgtgatctgttctggactctcgagtttttag
GATA2_c.1017+572C>T_R	ctaaaaactcgagagtcagaaacagatacacgaagtt
GATA2_c.1017+532T>A_F	agtttcctatccggacaactgcagcctgaagataa
GATA2_c.1017+532T>A_R	ttatcttcaggctgcagttgtccggataggaaact
GATA2_c.1017+582G>T_F	ctgtttccggactctgctagtttttagagtctcct
GATA2_c.1017+582G>T_R	aggagactctaaaaactagcagagtcggaaacag
GATA2_c.1017+693dup_F	aataatgcgtaaaatgctaagaataataatatttttttcaggcggaagtgattatg
GATA2_c.1017+693dup_R	cataaatcacttcgacctgaaaaaataattttattattcttagcttttacgcattatt
GATA2_c.1061C>T(p.T354M)_F	gttgtgcaaattgtcagatgacaaccaccaccttatg
GATA2_c.1061C>T(p.T354M)_R	cataaggtggtggtgtcatctgacaattgcacaac

Las secuencias de los oligonucleótidos se presentan en dirección 5'-3'. Los cebadores diseñados para el kit *QuikChange Lightning Site-Directed Mutagenesis* (Agilent) se basan en la auto-complementariedad de los mismos y presentan la región mutante en el centro del *primer*. La longitud de los cebadores es determinante en este protocolo, ya que de ella depende la cantidad de cebadores que se añade a cada reacción de PCR de mutagénesis para alcanzar una masa constante de 125 ng de cada oligonucleótido por reacción. Aunque los cebadores reciben la nomenclatura de fwd y rev (F y R), el producto de PCR abarca la extensión completa del ADN molde (plásmido silvestre).

Tabla Suplementaria 3. Rendimiento de la extracción de ARN de los cultivos transfectados con el vector silvestre.

Sample	[Nucleic acid] (ng/ μ L)	A ₂₆₀	A ₂₈₀	Ratio 260/280	Ratio 260/230	Type
GATA2_wt_11kb_M1	1296,4	32,41	15,356	2,11	2,03	RNA
GATA2_wt_11kb_M2	555,9	13,899	6,589	2,11	0,97	RNA
GATA2_wt_11kb_M3	759,4	18,985	8,98	2,11	2,06	RNA
RNA HEK293 NT	650,7	16,267	7,765	2,09	1,84	RNA

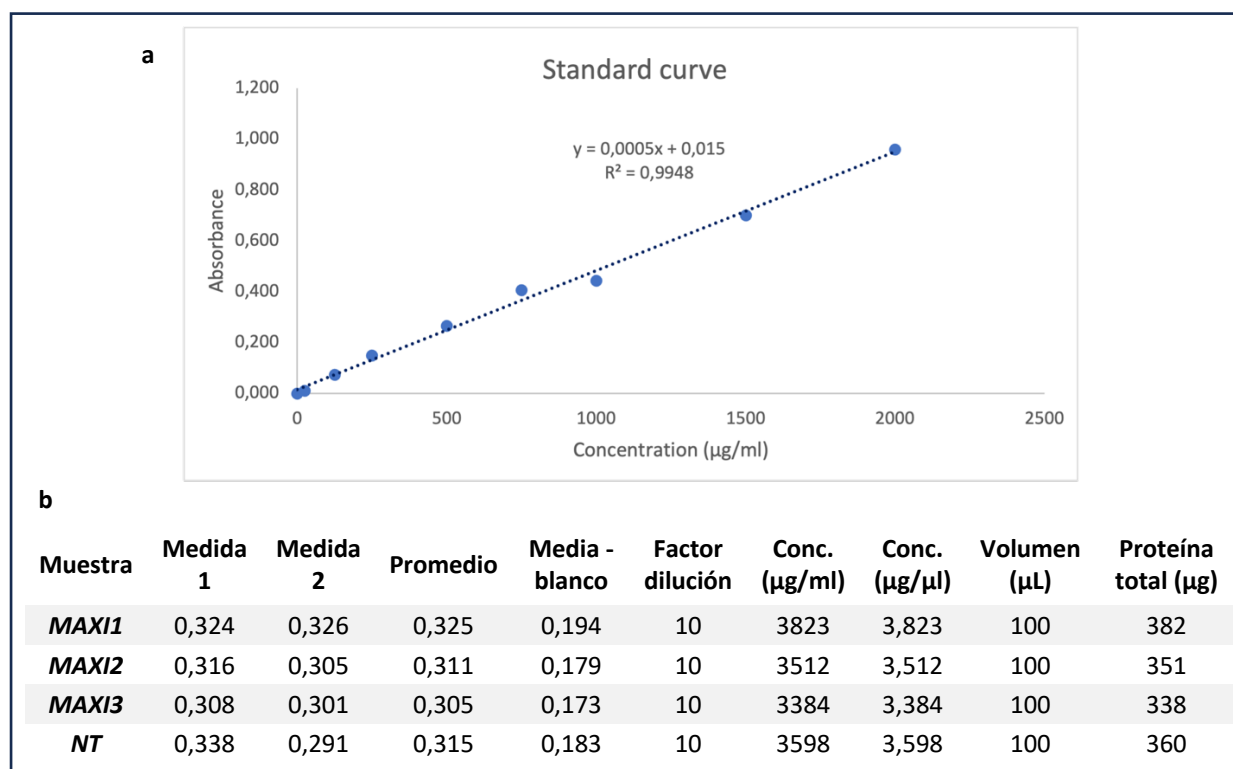


Figura Suplementaria 2. Cuantificación de la extracción de proteínas de los cultivos transfectados con el vector silvestre. **a** Curva patrón generada a partir de disoluciones de concentración conocida (0-2000 μ g/ml) de la proteína BSA. Se presenta la bondad del ajuste ($R^2=0,995$) **b** Cálculo de la concentración de proteína en las muestras problema a partir de la interpolación con la recta patrón.

Tabla Suplementaria 5. Resultado de la cuantificación de la densidad celular y de la fluorescencia mediante ImageJ de los cultivos transfectados con el plásmido silvestre

	Fondo	CC	CV	Confluencia (%)
WT_1	3145728	1658776	424205	52,7
WT_1	3145728	1153566	103918	36,7
WT_2	3145728	1235798	99820	39,3
WT_2	3145728	1286188	137188	40,9
WT_3	3145728	1432190	136457	45,5
WT_3	3145728	931219	64602	29,6

El valor de fondo corresponde al número de píxeles de cada imagen, siendo su resolución de 2048x1536 píxeles. La confluencia celular se estimó como el cociente entre CC y el Fondo, que corresponde al área celular ocupada.

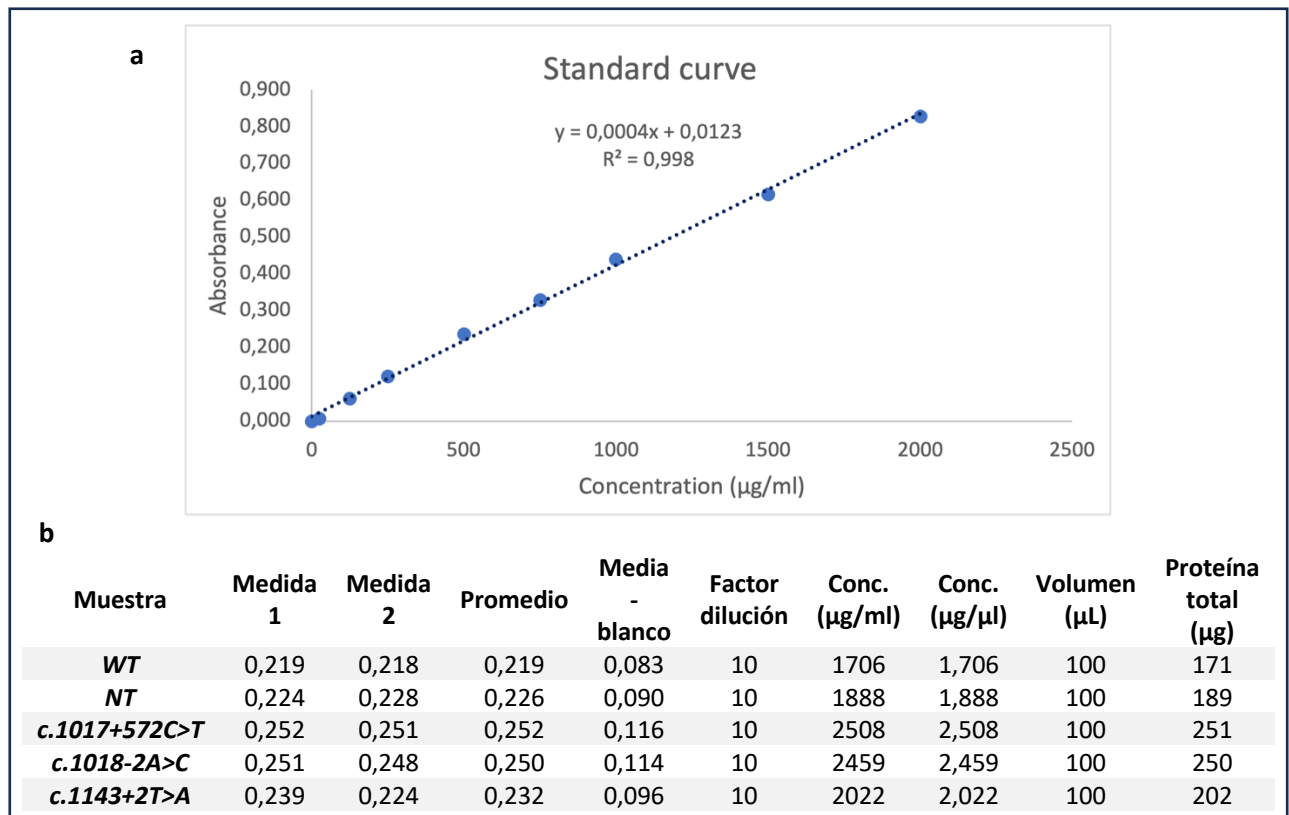


Figura Suplementaria 3. Cuantificación de la extracción de proteínas de los cultivos transfectados con los vectores mutantes. **a** Curva patrón generada a partir de disoluciones de concentración conocida (0-2000 µg/mL) de la proteína BSA. Se presenta la bondad del ajuste ($R^2=0,998$) **b** Cálculo de la concentración de proteína en las muestras problema a partir de la interpolación con la recta patrón.

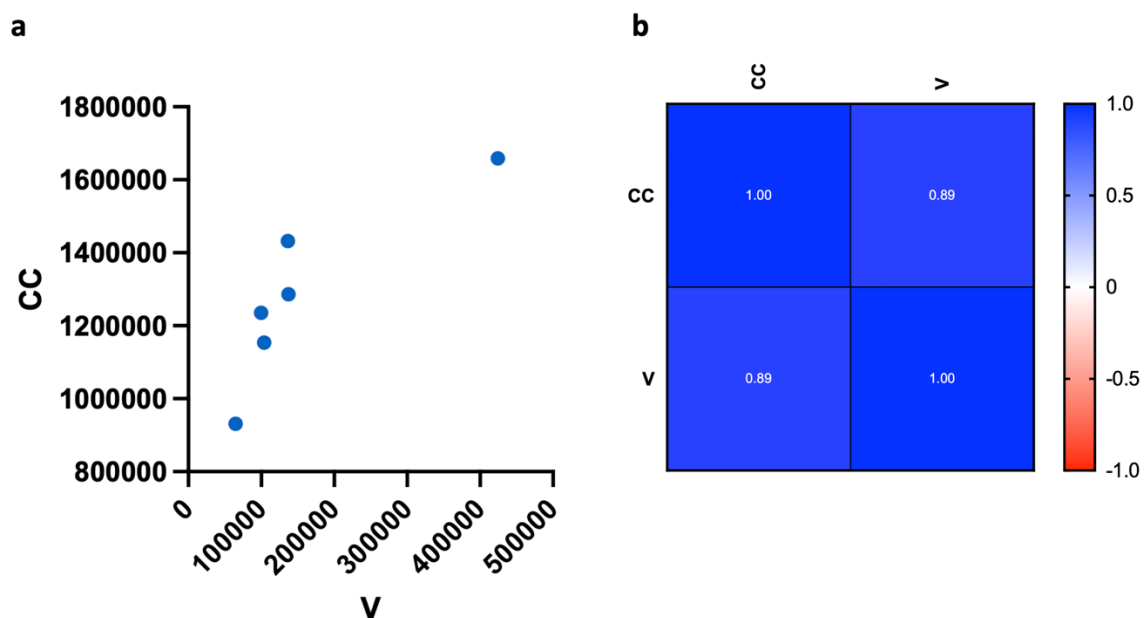


Figura Suplementaria 4. a Representación de la densidad celular (CC) frente a la fluorescencia verde (CV). Datos obtenidos a partir de la cuantificación realizada con ImageJ (píxeles). **b Matriz de correlación entre ambas variables.** Test de correlación de Spearman con un resultado de $r=0,8857$ y un p-valor asociado de $0,033$ (*) $n = 6$ (G.L.: 5)

Tabla Suplementaria 6. Resultado de la cuantificación de la densidad celular y de la fluorescencia mediante ImageJ de los cultivos transfectados con las variantes

<i>Transfección</i>	Fondo	CC	CV	Confluencia (%)	Ratio CV/CC (base 1)
<i>WT</i>	1572864	1102828	98992	70,1	0,08976196
	1572864	1041631	112105	66,2	0,10762449
	1572864	767290	100148	48,8	0,13052171
	1572864	818959	102753	52,1	0,12546782
	1572864	1118728	78870	71,1	0,07049971
	1572864	1056961	67616	67,2	0,06397209
	1572864	739092	52177	47,0	0,07059608
	1572864	702616	60587	44,7	0,08623060
<i>c.1017+572C>T</i>	786432	436792	23011	55,5	0,05268183
	786432	304588	11619	38,7	0,03814661
	786432	323860	16972	41,2	0,05240536
	786432	530232	18287	67,4	0,03448868
	786432	437526	15459	55,6	0,03533276
	786432	314144	14206	39,9	0,04522130
	786432	617349	26974	78,5	0,04369328
	786432	218644	7904	27,8	0,03615009
<i>c.1018-2A>C</i>	786432	574972	17699	73,1	0,03078237
	786432	441387	13986	56,1	0,03168648
	786432	395507	10958	50,3	0,02770621
	786432	646154	15989	82,2	0,02474488
	786432	617604	19393	78,5	0,03140038
	786432	343848	9497	43,7	0,02761976
	786432	457355	10880	58,2	0,02378896
	786432	559293	11528	71,1	0,02061174
<i>c.1143+2T>A</i>	786432	485029	17943	61,7	0,03699366
	786432	633102	25737	80,5	0,04065222
	786432	550054	14270	69,9	0,02594291
	786432	518722	15676	66,0	0,03022043
	786432	473750	13170	60,2	0,02779947
	786432	417268	10781	53,1	0,02583711
	786432	580865	17595	73,9	0,03029103
	786432	356318	13590	45,3	0,03814009

Para el cálculo de las áreas correspondientes al campo claro (CC) y campo verde (CV) se indicaron al software unos valores de corte de 90 y 25. Estos valores de corte maximizaron el área celular y fluorescente detectada sin sobrequantificar dichas variables. El valor de fondo corresponde al número de píxeles de cada imagen, siendo su resolución de 2048x1536 píxeles. La confluencia celular se estimó como el cociente entre los valores de CC y Fondo, que corresponde al área celular ocupada en la sección.

Tabla Suplementaria 7. Rendimiento de la extracción de ARN de los cultivos transfectados con los vectores mutantes

Sample	[Nucleic acid] (ng/ μ L)	A ₂₆₀	A ₂₈₀	Ratio 260/280	Ratio 260/230	Type
WT	741,8	18,545	8,878	2,09	2,16	RNA
c.1017+572C>T	654,9	16,373	7,731	2,12	1,28	RNA
c.1018-2A>C	558,7	13,968	6,606	2,11	1,24	RNA
c.1143+2T>A	515,9	12,879	6,273	2,05	1,81	RNA
NT	456,8	11,421	5,625	2,03	1,15	RNA

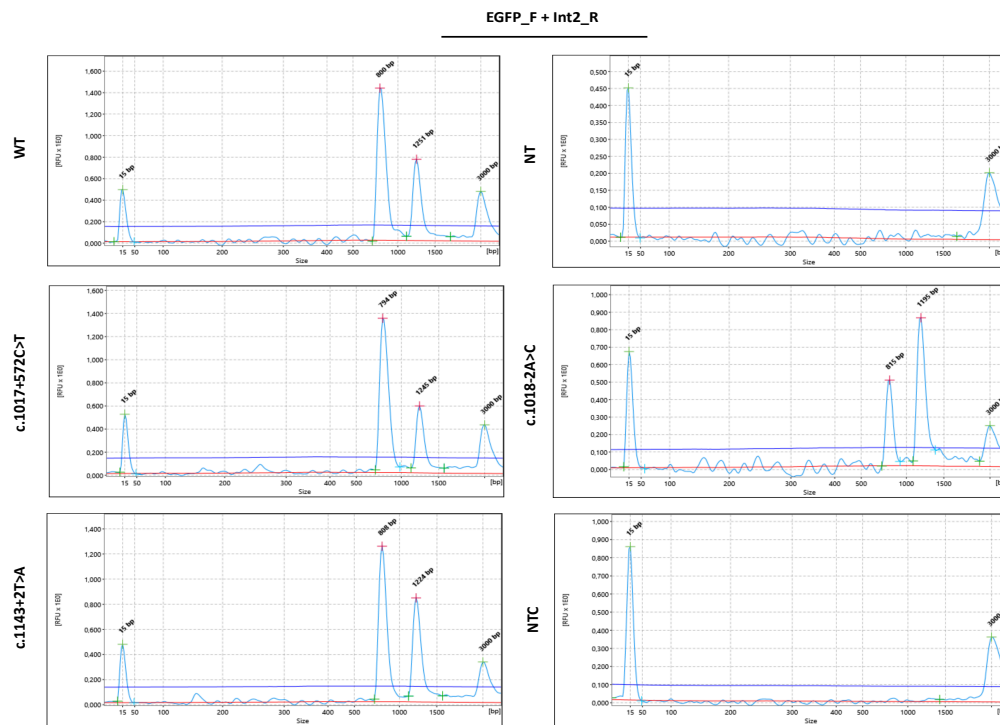


Figura Suplementaria 5. Electroferogramas correspondientes a la amplificación por PCR con cebadores específicos del vector (EGFP_F/Int2_R).

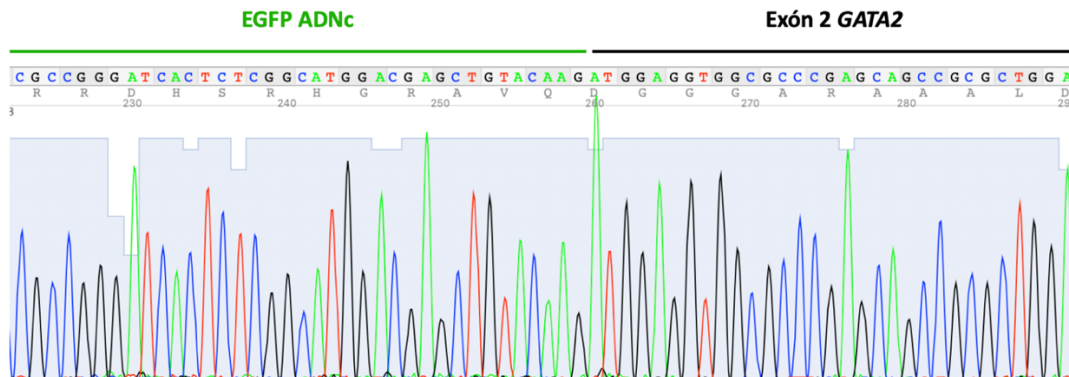


Figura Suplementaria 6. Electroferograma correspondiente a la secuenciación mediante Sanger con el cebador específico del vector EGFP_F. Se muestra la región que contiene la ADNc de EGFP como marca específica de la expresión del ORF del vector.

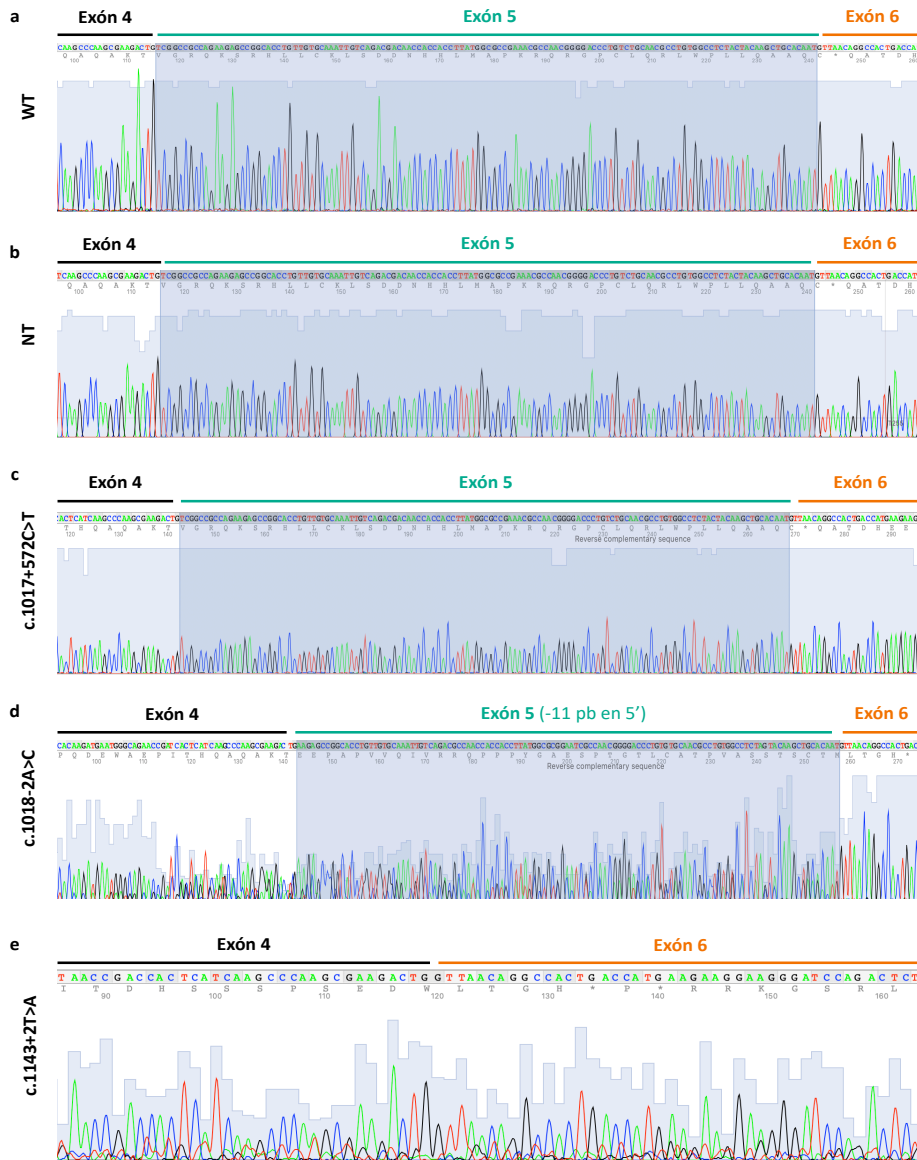
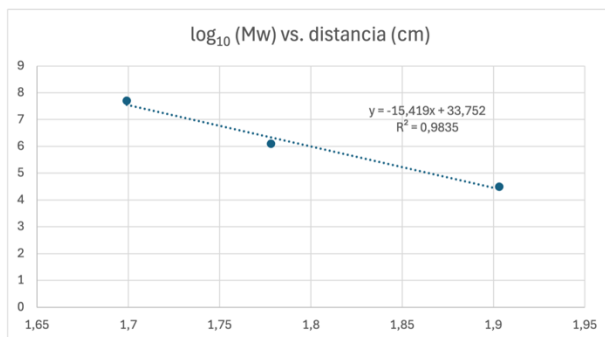


Figura Suplementaria 7. Electroferogramas correspondientes a la secuenciación mediante Sanger del ADNc extraído de cada cultivo celular. **a** Lectura con el cebador E46_F sobre el ADNc del cultivo transfectado con el vector con la secuencia silvestre de GATA2 **b** Lectura con E46_F sobre el ADNc del cultivo no transfectado **c, d, e** Lecturas con los cebadores E46_F, E46_R y E46_F del ADNc con los cultivos transfectados con los vectores mutantes c.1017+572C>T, c.1018-2A>C y c.1143+2T>A, respectivamente. La calidad de la secuencia es baja en aquellas reacciones en las que hay más de un amplicón diferente.



Muestra	Problema	Interpolación	Masa (kDa)
WT	4,6	1,891	77,7
c.1143+2T>A	5	1,865	73,2
c.1018-2A>C	5,5	1,832	68,0
c.1017+572C>T	5,9	1,806	64,0
	4,65	1,887	77,2

Figura Suplementaria 8. Estimación de la masa proteica de las bandas observadas en el Western blot de los extractos proteicos de los cultivos celulares. La distancia recorrida por cada banda se determinó manualmente con una regla de laboratorio.

Secuencia silvestre de ADNc de GATA2 (480 aminoácidos, 79,7 kDa)

MEVAPEQPRWMAHPAVLNAQHPDSHHPGLAHNYMEPAQLLPPDEVDFVFNHLD SQGNPYYPANPAHARARV
 SYSPAHARLTGGQMCRPHLLHSPGLPWLDGGKAALSAAAAHHHNPWTVSPFSKTPHPSAAGGPGGGLSVYP
 GAGGGSGGGSGSSVASLTPTAAHSGSHLFGFPPTPPKEVSPDPSTTGAASPASSSAGGSAARGEDKDGVKYQVS
 LTESMKMESGSLRPGLATMGTQPATHHPIPTYPSYVAAAHDYSSGLFHPGGFLGGPASSFTPKQRSKARSCSE
 GREVCNCGATATPLWRRDGTGHYLCNACGLYHKMNGQNRPLIKPKRRLSAARRAGTCCANCQTTTTTLWRRN
ANGDPVCNACGLYKLNHNVRPLTMKKEGIQTRNRKMSNKS KSKKGAECFEELSKCMQEKSPPFSAALAGH
MAPVGHLPFFSHSGHILPTPTPIHPSSLSFGHHPSSMVTAMG

c.1018-2A>C: Creación de un sitio críptico en el exón 5 que produce la pérdida de 11pb del extremos 5' de este exón (378 aminoácidos, 68,48 kDa)

MEVAPEQPRWMAHPAVLNAQHPDSHHPGLAHNYMEPAQLLPPDEVDFVFNHLD SQGNPYYPANPAHARARV
 SYSPAHARLTGGQMCRPHLLHSPGLPWLDGGKAALSAAAAHHHNPWTVSPFSKTPHPSAAGGPGGGLSVYP
 GAGGGSGGGSGSSVASLTPTAAHSGSHLFGFPPTPPKEVSPDPSTTGAASPASSSAGGSAARGEDKDGVKYQVS
 LTESMKMESGSLRPGLATMGTQPATHHPIPTYPSYVAAAHDYSSGLFHPGGFLGGPASSFTPKQRSKARSCSE
 GREVCNCGATATPLWRRDGTGHYLCNACGLYHKMNGQNRPLIKPKRRLKSRHLLCKLSDDNHHLMAPKRQRG
PCLQRLWPLLQAAQC

c.1018-2A>C: Creación de un sitio críptico aceptor en el intrón que produce retención intrónica parcial (346 aminoácidos, 64,96 kDa):

MEVAPEQPRWMAHPAVLNAQHPDSHHPGLAHNYMEPAQLLPPDEVDFVFNHLD SQGNPYYPANPAHARARV
 SYSPAHARLTGGQMCRPHLLHSPGLPWLDGGKAALSAAAAHHHNPWTVSPFSKTPHPSAAGGPGGGLSVYP
 GAGGGSGGGSGSSVASLTPTAAHSGSHLFGFPPTPPKEVSPDPSTTGAASPASSSAGGSAARGEDKDGVKYQVS
 LTESMKMESGSLRPGLATMGTQPATHHPIPTYPSYVAAAHDYSSGLFHPGGFLGGPASSFTPKQRSKARSCSE
 GREVCNCGATATPLWRRDGTGHYLCNACGLYHKMNGQNRPLIKPKRRLGEVIYEA

Acceptor site predictions for 10.42.3.121.1370294.0:

Start	End	Score	Intron	Exon
145	185	0.92	gtcgtctctctctttctggcc aG attctttcgggccagatttc	
179	219	0.82	agatttcctcctcgggtat aG ggagcccaccgggcaccctt	
277	317	0.83	tgtcctctggcctctgctt aG ctcttttttaaaaatagggc	
681	721	0.95	ataataaatatattttttc aG ggccaagtgatttatgagggc	
882	922	0.79	ctctgtcccttctgtgacc aG gcttgggcctggggctgttc	
914	954	0.55	gggctgttccaggetctgc aG gcctcagccccagcccccc	
966	1006	0.70	ctgggtgcactcccgcctgc aG ttctctgggaagtgttgggg	
1116	1156	0.43	tccagaccctacctctgcc aG gcatttcctgagggactagg	
1349	1389	0.88	tttaattttgttaatttgt aG ttagatcttattttaatggc	
1580	1620	0.51	ctactagagagcccttctc aG ggcctcagtttccccattta	
1651	1691	0.91	gtagctaattctcctctgt aG ctcttgcaatcccgttgatt	

A las masas moleculares de las proteínas se les ha sumado la masa individual de la EGFP para realizar la comparación con las bandas del *Western blot*. Asimismo, se incorporan las predicciones de NNSPLICE en el intrón 4 de GATA2 y se ha marcado en negrita la opción más probable (0,95). En las secuencias proteicas, las regiones diferentes entre las 3 isoformas se han subrayado.