



**UNIVERSIDAD POLITÉCNICA SALESIANA
SEDE GUAYAQUIL**

CARRERA DE INGENIERÍA EN BIOTECNOLOGÍA

DISEÑO *IN SILICO* DE UN SISTEMA CRISPR INTERFERENCIA

DIRIGIDO A INHIBIR LA EXPRESIÓN DEL

ONCOGÉN *HOXB13* EN CÁNCER DE PRÓSTATA

*Trabajo de titulación previo a la obtención
del título de Ingeniero en Biotecnología*

AUTOR:

RICHARD STEVEN MARCILLO ZAMBRANO

TUTOR:

Mgtr. CRISTIAN ALEXANDER ZAMBRANO MENDOZA

GUAYAQUIL - ECUADOR

2026

**CERTIFICADO DE RESPONSABILIDAD Y AUTORÍA DEL TRABAJO DE
TITULACIÓN**

Yo, **Richard Steven Marcillo Zambrano** con documento de identificación N° 0954989422;
Manifiesto que:

Soy el autor y responsable del presente trabajo; y, autorizo a que sin fines de lucro la Universidad Politécnica Salesiana pueda usar, difundir, reproducir o publicar de manera total o parcial el presente trabajo de titulación.

Guayaquil, 02 de febrero del año 2026

Atentamente,

Richard Marcillo

Richard Steven Marcillo Zambrano

CI: 0954989422

**CERTIFICADO DE CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE
TITULACIÓN A LA UNIVERSIDAD POLITÉCNICA SALESIANA**

Yo, **Richard Steven Marcillo Zambrano** con documento de identificación N° 0954989422, expresado mi voluntad y por medio del presente documento cedo a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que soy autor/a del trabajo experimenta: **DISEÑO *IN SILICO* DE UN SISTEMA CRISPR INTERFERENCIA DIRIGIDO A INHIBIR LA EXPRESIÓN DEL ONCOGÉN *HOXB13* EN CÁNCER DE PRÓSTATA**, el cual ha sido desarrollado para optar por el título de: *Ingeniero en Biotecnología*, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En concordancia con lo manifestado, suscribimos este documento en el momento que hacemos la entrega del trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

Guayaquil, 02 de febrero del año 2026

Atentamente,

Richard Marcillo

Richard Steven Marcillo Zambrano

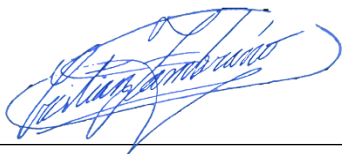
CI: 0954989422

CERTIFICADO DE DIRECCIÓN DEL TRABAJO DE TITULACIÓN

Yo, **Cristian Alexander Zambrano Mendoza** con documento de identificación N° 0930726112, docente de la Universidad Politécnica Salesiana, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: **DISEÑO *IN SILICO* DE UN SISTEMA CRISPR INTERFERENCIA DIRIGIDO A INHIBIR LA EXPRESIÓN DEL ONCOGÉN *HOXB13* EN CÁNCER DE PRÓSTATA**, realizado por **Richard Steven Marcillo Zambrano** con documento de identificación N° 0954989422, obteniendo como resultado final el trabajo de titulación bajo la opción Trabajo Experimental que cumple con todos los requisitos determinados por la Universidad Politécnica Salesiana.

Guayaquil, 02 de febrero del año 2026

Atentamente,



Cristian Alexander Zambrano Mendoza

CI: 0930726112

CERTIFICADO DE DIRECCIÓN DEL TRABAJO DE TITULACIÓN

Yo, **Brenda Patricia López Ulloa** con documento de identificación N° 0960450922, docente de la Universidad Politécnica Salesiana, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: **DISEÑO *IN SILICO* DE UN SISTEMA CRISPR INTERFERENCIA DIRIGIDO A INHIBIR LA EXPRESIÓN DEL ONCOGÉN *HOXB13* EN CÁNCER DE PRÓSTATA**, realizado por **Richard Steven Marcillo Zambrano** con documento de identificación N° 0954989422, obteniendo como resultado final el trabajo de titulación bajo la opción Trabajo Experimental que cumple con todos los requisitos determinados por la Universidad Politécnica Salesiana.

Guayaquil, 02 de febrero del año 2026

Atentamente,



Brenda Patricia López Ulloa

CI: 0960450922

Dedicatoria

A mis padres, mami Ruth y Richard, quienes siempre me han brindado su apoyo en mis elecciones, y han hecho todo lo que está a su alcance para que yo cumpliera lo que me proponga.

A mi abuela, mi mami tere, por su cuidado incondicional, su cariño diario y el amor sincero con el que siempre me ha acompañado. Su calidez ha sido refugio en cada etapa de mi vida.

A mis padrinos, tío Ronald y tía Liz por su apoyo permanente y por enseñarme, con palabras y acciones, el valor de elegir siempre el camino correcto.

A mis tíos y primas de Manta, quienes siempre me han hecho sentir un hogar en cualquier lugar junto a ellos.

A mis compañeros, quienes terminaron integrándose en mi ADN: Sasa, José, Diego, Katherine, Britany, Ashley.

Agradecimientos

A la Universidad Politécnica Salesiana, por brindarme los espacios, la formación y los años de aprendizaje que hicieron posible mi desarrollo académico y personal a lo largo de la carrera.

A mi tutor, Ing. Cristian Zambrano, por asumir con compromiso y responsabilidad la guía de este trabajo durante el período de integración curricular. Su acompañamiento constante, su criterio técnico y su alto nivel de exigencia fueron determinantes para elevar el rigor académico de esta investigación. A través de sus observaciones y estándares de trabajo, fortalecí mi capacidad crítica y mi forma de abordar problemas científicos, aprendizajes que sin duda tendrán un impacto directo en mi desarrollo profesional más allá de este proyecto.

A mi co-tutora, Dra. Brenda López por su enseñanza a lo largo de la carrera y por despertar en mí un interés profundo por la biología molecular y sintética. Por ser una fuente permanente de inspiración para aspirar a metas más altas y asumir nuevos retos con convicción.

Al Ing. Jaime Naranjo por su disposición permanente para escuchar, orientar y enseñar. Por ser una puerta siempre abierta y un apoyo en momentos clave del proceso formativo.

A la Ing. Elizabeth Marcillo por su apoyo incondicional, por orientarme con criterio y sensibilidad en la elección de esta carrera a partir de mis intereses y aptitudes, por motivarme con su ejemplo a seguir el camino de la ciencia y por enseñarme el valor de la trascendencia académica y humana.

Resumen

El cáncer de próstata constituye una de las principales neoplasias malignas en la población masculina y presenta una fracción relevante de casos con componente hereditario. Entre los genes asociados a este tipo de cáncer, *HOXB13* ha sido identificado como un regulador central en la progresión tumoral prostática, con funciones transcripcionales y epigenéticas críticas. No obstante, su naturaleza como factor de transcripción lo clasifica como una diana farmacológicamente no drogable, lo que limita el abordaje mediante estrategias terapéuticas convencionales. En este contexto, la interferencia mediada por CRISPR (*CRISPRi*) emerge como una alternativa para el silenciamiento transcripcional específico sin inducir roturas en el ADN. El objetivo de esta investigación fue diseñar *in silico* un sistema *CRISPRi* dirigido a inhibir la expresión del oncogén *HOXB13* en cáncer de próstata. Para ello, se desarrolló un flujo metodológico computacional estructurado en tres fases. En la Fase 1 se caracterizó el locus genómico de *HOXB13* en el genoma humano (GRCh38/hg38) y se delimitó una región reguladora proximal al sitio de inicio de la transcripción. En la Fase 2 se diseñaron y evaluaron guías de ARN mediante las plataformas CHOPCHOP, CRISPOR, Cas-OFFinder y RNAfold, integrando criterios de posición relativa al TSS, especificidad genómica, análisis exhaustivo de off-targets y consistencia estructural del ARN guía. En la Fase 3 se realizó el diseño *in silico* de constructos plasmídicos *CRISPRi* derivados de un vector lentiviral validado, incorporando las guías seleccionadas sin modificar la arquitectura del backbone. Los resultados permitieron priorizar cinco crRNAs con perfiles compatibles con interferencia transcripcional, ausencia de off-targets críticos y estructuras secundarias funcionales. El análisis estadístico exploratorio indicó un conjunto homogéneo entre candidatas, sin diferencias globales marcadas, lo que justificó la integración jerárquica de múltiples parámetros para su priorización racional. El diseño plasmídico final integra de forma coherente el cassette hU6-sgRNA y el efector dCas9-KRAB, constituyendo una plataforma consistente para estudios posteriores. En conjunto, este trabajo establece un diseño computacional reproducible de un sistema *CRISPRi* dirigido a *HOXB13*, que sienta las bases metodológicas para su validación experimental futura en modelos de cáncer de próstata.

Palabras clave: cáncer de próstata; *CRISPRi*; *HOXB13*; diseño *in silico*; silenciamiento génico

Abstract

Prostate cancer constitutes one of the major malignant neoplasms in the male population and presents a relevant fraction of cases with a hereditary component. Among the genes associated with this type of cancer, *HOXB13* has been identified as a central regulator in prostate tumor progression, with critical transcriptional and epigenetic functions. However, its nature as a transcription factor classifies it as a pharmacologically non-druggable target, limiting conventional therapeutic strategies. In this context, CRISPR-mediated interference (CRISPRi) emerges as an alternative for specific transcriptional silencing without inducing DNA double-strand breaks.

The objective of this research was to design *in silico* a CRISPRi system aimed at inhibiting the expression of the *HOXB13* oncogene in prostate cancer. To this end, a computational methodological workflow structured in three phases was developed. In Phase 1, the *HOXB13* genomic locus was characterized in the human genome (GRCh38/hg38), and a regulatory region proximal to the transcription start site was delimited. In Phase 2, guide RNAs were designed and evaluated using the CHOPCHOP, CRISPOR, Cas-OFFinder, and RNAfold platforms, integrating criteria of relative position to the TSS, genomic specificity, exhaustive off-target analysis, and guide RNA structural consistency. In Phase 3, *in silico* design of CRISPRi plasmid constructs derived from a validated lentiviral vector was performed, incorporating the selected guides without modifying the backbone architecture.

The results allowed the prioritization of five crRNAs with profiles compatible with transcriptional interference, absence of critical off-targets, and functional secondary structures. Exploratory statistical analysis indicated a homogeneous set among candidates, without marked global differences, which justified the hierarchical integration of multiple parameters for rational prioritization. The final plasmid design coherently integrates the hU6–sgRNA cassette and the dCas9-KRAB effector, constituting a consistent platform for subsequent studies. Overall, this work establishes a reproducible computational design of a CRISPRi system targeting *HOXB13*, laying the methodological foundations for future experimental validation in prostate cancer models.

Keywords: prostate cancer; CRISPRi; *HOXB13*; *in silico* design; gene silencing

Índice de Contenido

CAPÍTULO 1	1
Antecedentes	1
1.1. Antecedentes	1
1.2. Introducción	3
1.3. Planteamiento del problema	5
1.4. Justificación.....	6
1.5. Objetivos	9
1.5.1. Objetivo general	9
1.5.2. Objetivos específicos.....	9
1.6. Hipótesis.....	9
CAPÍTULO 2	10
Marco teórico	10
2.1. Los genes <i>HOX</i> como reguladores del desarrollo y su implicación en cáncer..	10
2.2. <i>HOXB13</i> como regulador crítico en células de cáncer de próstata	10

2.3. <i>HOXB13</i> como diana transcripcional y limitaciones farmacológicas.....	12
2.4. Interferencia CRISPR como estrategia de silenciamiento transcripcional.....	12
2.5. Justificación del diseño <i>in silico</i> del sistema CRISPRi contra <i>HOXB13</i>	13
CAPÍTULO 3.....	15
Metodología.....	15
3.1. Fase 1: Caracterización genómica del gen	15
3.1.1. Identificación del gen <i>HOXB13</i> y extracción de la región diana en UCSC Genome Browser.....	15
3.2. Fase 2: Diseño y preselección bioinformática de ARN guía (sgRNAs) dirigidos a <i>HOXB13</i>	17
3.2.1. Generación y preselección de crRNAs candidatos mediante CHOPCHOP	17
3.2.2. Determinación de coordenadas genómicas y cálculo de distancia al TSS para crRNAs candidatas.....	19
3.2.3. Evaluación de especificidad genómica y depuración avanzada de crRNAs mediante CRISPOR.....	20
3.2.4. Depuración exhaustiva de off-targets genómicos mediante Cas-OFFinder22	
3.2.5. Evaluación de la consistencia estructural del sgRNA mediante RNAfold	23
3.2.6. Integración de resultados de la Fase 2.....	24

3.2.7. Análisis estadístico exploratorio y priorización multicriterio de crRNAs .	25
3.2.7.1. Enfoque general del análisis estadístico.....	25
3.2.7.2. Preparación y estructuración de los datos	26
3.2.7.3. Análisis comparativo no paramétrico.....	26
3.2.7. 4. Análisis de correlación entre variables.....	26
3.2.7.5. Normalización de variables y análisis multicriterio	27
3.3. Fase 3: Diseño <i>in silico</i> del sistema CRISPRi y generación de constructos plasmídicos.....	27
3.3.1. Selección del vector base para la implementación del sistema CRISPRi ..	27
3.3.2. Arquitectura general del vector y orientación de los cassettes de expresión	28
3.3.3. Inserción <i>in silico</i> del espaciador del sgRNA en el cassette hU6	28
3.3.4. Conservación del cassette efector dCas9-KRAB y del marcador fluorescente	29
3.3.5. Documentación esquemática del diseño plasmídico	29
3.3.6. Alcance de la Fase 3.....	31
CAPÍTULO 4.....	32
Resultados y discusión	32

4.1.1. Resultados – Fase 1: Caracterización genómica del gen <i>HOXB13</i>	32
4.1.1.1. Análisis en USCS Genome Browser	32
4.1.2. Resultados-Fase 2: Diseño y preselección bioinformática de ARN guía (sgRNAs) dirigidos a <i>HOXB13</i>	32
4.1.2.1. Generación y preselección de crRNAs candidatos mediante CHOPCHOP	32
4.1.2.2. Determinación de coordenadas genómicas y posición relativa al TSS de las crRNAs preseleccionadas	34
4.1.2.3. Evaluación de especificidad genómica de crRNAs mediante CRISPOR35	
4.1.2.4. Evaluación exhaustiva de especificidad genómica de crRNAs mediante Cas-OFFinder	36
4.1.2.5. Evaluación de la consistencia estructural del sgRNA mediante RNAfold	37
4.1.2.5.1. Análisis cuantitativo comparativo	38
4.1.2.5.2. sgRNA Rank 1	39
4.1.2.5.3 sgRNA Rank 3	40
4.1.2.5.4 sgRNA Rank 6	42
4.1.2.5.5 sgRNA Rank 10	44
4.1.2.5.6 sgRNA Rank 12	46

4.1.2.6. Integración de resultados de la Fase 2.....	48
4.1.2.7. Resultados del análisis estadístico de la Fase 2.....	51
4.1.2.7.1. Comparación de estabilidad estructural por MFE (RNAfold)	51
4.1.2.7.2. Asociación entre puntajes de especificidad MIT y CFD (CRISPOR) .	52
4.1.2.7.3. Correlaciones entre parámetros estructurales y genómicos	52
4.1.2.7.4 Ranking multicriterio de crRNAs	53
4.1.3. Resultados de la Fase 3: Generación y validación in silico de constructos CRISPRi.....	54
4.1.3.1. Generación de constructos in silico derivados del vector base	54
4.1.3.2. Documentación esquemática del constructo representativo.....	55
4.2. Discusión.....	57
5.1. Conclusiones	62
Conclusión 1 – Objetivo específico 1	62
Conclusión 2 – Objetivo específico 2	62
Conclusión 3 – Objetivo específico 3	62
Conclusión general.....	63

5.2. Recomendaciones.....	63
Referencias Bibliográficas	64
Anexos.....	72

Abreviaturas

AR	Receptor de andrógenos
AR-V7	Variante de empalme 7 del receptor de andrógenos
ATAC-seq	Secuenciación de accesibilidad cromatínica mediante transposasa
Cas9	Proteína asociada a CRISPR 9
CDS	Secuencia codificante
CFD	Determinación de la frecuencia de corte
CHOPCHOP	Plataforma bioinformática para diseño de guías CRISPR
CRISPOR	Plataforma para análisis de especificidad y eficiencia CRISPR
CRISPR	Repeticiones palindrómicas cortas agrupadas y regularmente interespaciadas
CRISPRi	Interferencia CRISPR
dCas9	Cas9 catalíticamente inactiva
DNA	Ácido desoxirribonucleico
DNase-seq	Secuenciación de sitios hipersensibles a DNasa I
EGFP	Proteína verde fluorescente mejorada

FASTA	Formato de secuencia biológica
GC	Contenido de guanina y citosina
GFP	Proteína verde fluorescente
GRCh38	Ensamblaje de referencia del genoma humano 38
HDAC3	Histona desacetilasa 3
hg38	Versión 38 del genoma humano
KRAB	Caja asociada a Krüppel
LNCaP	Línea celular humana de cáncer de próstata
MIT	Instituto Tecnológico de Massachusetts
MFE	Energía mínima libre
NCoR	Correpressor nuclear del receptor
NCCN	Red Nacional Integral del Cáncer
NLS	Señal de localización nuclear
OT	Fuera de objetivo (off-target)
PAM	Motivo adyacente al protoespaciador
pb	Pares de bases

pf	Función de partición
RefSeq	Base de datos de secuencias de referencia
rCREs	Elementos reguladores <i>cis</i> asociados a riesgo
RNA	Ácido ribonucleico
sgRNA	ARN guía único
siRNA	ARN interferente pequeño
SMRT	Mediador de silenciamiento para receptores de retinoides y tiroides
SpCas9	Cas9 derivada de <i>Streptococcus pyogenes</i>
T2A	Péptido autoclivable 2A
TSS	Sitio de inicio de la transcripción
UCSC	Universidad de California, Santa Cruz

Simbología

°C	Grados Celsius
%	Porcentaje
-	Orientación negativa / valor negativo
+	Orientación positiva / valor positivo
≤	Menor o igual que
≥	Mayor o igual que
~	Aproximadamente
5'	Extremo cinco prima
3'	Extremo tres prima
bp	Pares de bases
kcal/mol	Kilocalorías por mol
mm	Desajustes (mismatches)
nt	Nucleótidos
→	Dirección
±	Más o menos

Índice de figuras

Figura 1. Sistema CRISPRi.....	13
Figura 2. Arquitectura del vector pLV hU6-sgRNA hUbc-dCas9-KRAB-T2a-GFP utilizada como base para el diseño <i>in silico</i> de los constructos CRISPRi.....	30
Figura 3. Mountain plot y perfil de entropía del sgRNA Rank 1.....	39
Figura 4. Análisis estructural integrado del sgRNA Rank 1 mediante RNAfold.	40
Figura 5. Mountain plot y perfil de entropía del sgRNA Rank 3.....	41
Figura 6. Análisis estructural integrado del sgRNA Rank 3 mediante RNAfold.	42
Figura 7. Mountain plot y perfil de entropía del sgRNA Rank 6.....	43
Figura 8. Análisis estructural integrado del sgRNA Rank 6 mediante RNAfold.	44
Figura 9. Mountain plot y perfil de entropía del sgRNA Rank 10.....	45
Figura 10. Análisis estructural integrado del sgRNA Rank 10 mediante RNAfold. ...	46
Figura 11. Mountain plot y perfil de entropía del sgRNA Rank 12.....	47
Figura 12. Análisis estructural integrado del sgRNA Rank 12 mediante RNAfold. ...	48
Figura 13. Comparación de estabilidad estructural por MFE.	51
Figura 14. Relación entre puntajes MIT y CFD.....	52

Figura 15. Ranking final de crRNAs por score integrado.....	53
Figura 16. Mapa de calor de parámetros normalizados por crRNA.....	54
Figura 17. Mapa plasmídico circular del constructo pLV hU6-sgRNA hUbc-dCas9-KRAB-T2a-GFP con inserción del espaciador de 20 nt correspondiente al Rank 12.	55
Figura 18. Mapa lineal del vector lentiviral pLV hU6-sgRNA hUbc-dCas9-KRAB-T2a-GFP (15 002 pb) generada <i>in silico</i> en SnapGene.	56

Índice de tablas

Tabla 1. Parámetros de caracterización genómica del gen <i>HOXB13</i> y definición de la región diana.	16
Tabla 2. Configuración general empleada en CHOPCHOP para el diseño de crRNAs dirigidas a <i>HOXB13</i>	17
Tabla 3. Características de las crRNAs seleccionadas mediante CHOPCHOP para CRISPRi del gen <i>HOXB13</i>	34
Tabla 4. Coordenadas genómicas y posición relativa al TSS de las crRNAs preseleccionadas mediante CHOPCHOP.....	35
Tabla 5. Resultados del análisis de especificidad genómica de crRNAs dirigidas al gen <i>HOXB13</i> mediante CRISPOR.....	36
Tabla 6. Perfil de especificidad genómica de crRNAs candidatos evaluados mediante Cas-OFFinder. ON = sitio en objetivo; OT = off-target; mm = mismatches.	37
Tabla 7. Parámetros termodinámicos de plegamiento secundario de sgrNAs candidatas obtenidos mediante RNAfold. MFE = energía mínima libre; Frec. MFE = frecuencia de la estructura MFE dentro del ensamble; Diversidad = diversidad conformacional del ensamble.	38
Tabla 8. Evaluación comparativa integral de cinco crRNAs candidatos para CRISPRi dirigido a <i>HOXB13</i>	50

Índice de anexos

Anexo 1. Salida Fase 1 de 351pb respecto al TSS	72
Anexo 2. Secuencia completa gen <i>HOXB13</i>	72
Anexo 3. Secuencia Scaffold.	74
Anexo 4. Información y secuencia del vector lentiviral CRISPRi pLV hU6-sgRNA hUbC-dCas9-KRAB-T2A-GFP.	74
Anexo 5. Secuencias de sgRNAs CRISPRi dirigidas a HOXB13 (crRNA + scaffold estándar)	79

CAPÍTULO 1

Antecedentes

1.1. Antecedentes

El cáncer de próstata constituye una de las principales neoplasias malignas en la población masculina y presenta una fracción relevante de casos con componente hereditario, en individuos con diagnóstico a edades tempranas o con antecedentes familiares positivos (Crawford et al., 2024; Ewing et al., 2012). En este contexto, la identificación de genes de susceptibilidad ha permitido avanzar en la comprensión de los determinantes genéticos asociados al riesgo de la enfermedad (Crawford et al., 2024; Ewing et al., 2012).

De acuerdo con las estimaciones de GLOBOCAN 2022, el cáncer de próstata se posiciona como la segunda causa de mortalidad por cáncer en hombres a nivel mundial, considerando todos los tipos de cáncer en población masculina (Bray et al., 2024). En Ecuador, los datos de mortalidad indican que el cáncer de próstata representa la principal causa de muerte por cáncer en hombres, con 1 362 defunciones estimadas, superando a otros tipos de cáncer frecuentes en el país (Bray et al., 2024).

En términos de incidencia, GLOBOCAN 2022 identifica al cáncer de próstata como el cáncer más diagnosticado en hombres a escala global, con más de 1,4 millones de nuevos casos nuevos (Bray et al., 2024). Este patrón se mantiene en América, Latinoamérica y Sudamérica, así como en Ecuador, donde se estimaron 3 456 nuevos casos en población masculina, posicionándolo como el primer cáncer en incidencia en hombres a nivel nacional (Bray et al., 2024).

Dentro de los genes vinculados al cáncer de próstata hereditario, el gen Homeobox B13 (*HOXB13*) ha sido identificado de como un factor de riesgo de alto impacto (Crawford et al., 2024; Dupont et al., 2021; Ewing et al., 2012). El estudio fundacional de Ewing et al. (2012) describió la mutación germinal recurrente c.251G>A:p.G84E como un determinante significativo de susceptibilidad al cáncer de próstata hereditario de inicio temprano. Esta asociación ha sido validada en estudios poblacionales de gran escala, donde se ha demostrado que la variante G84E confiere un riesgo aumentado no solo de cáncer de próstata, sino también

de enfermedad metastásica y mortal, como se evidenció en una cohorte de 592 158 hombres analizada por Crawford et al. (2024).

La relevancia clínica de la mutación G84E en *HOXB13* se ve reforzada por su incorporación en guías clínicas internacionales (Dupont et al., 2021). Dupont et al. (2021) reportaron que esta variante germinal confiere un riesgo clínico significativo y ha sido incluida en las guías del *National Comprehensive Cancer Network* (NCCN) como parte de los criterios de evaluación genética en cáncer de próstata hereditario, lo que consolida su importancia no solo desde una perspectiva epidemiológica, sino también en la práctica clínica.

Más allá de su asociación genética, *HOXB13* cumple un papel funcional central en la biología del cáncer de próstata (Lu et al., 2022; Xiang et al., 2025). Estudios recientes han demostrado que *HOXB13* actúa como un regulador transcripcional clave del receptor de andrógenos (AR) en cáncer de próstata avanzado, modulando la actividad de potenciadores y programas transcripcionales asociados a progresión tumoral y resistencia terapéutica (Xiang et al., 2025). De manera complementaria, se ha descrito que *HOXB13* ejerce funciones independientes del AR, actuando como regulador epigenético del metabolismo tumoral. En particular, Lu et al. (2022) demostraron que *HOXB13* recluta el complejo HDAC3–NCoR/SMRT para reprimir epigenéticamente genes lipogénicos, suprimiendo la lipogénesis *de novo* y modulando fenotipos tumorales asociados a agresividad.

La actividad oncogénica de *HOXB13* se encuentra regulada por mecanismos postraduccionales. Y. Chen et al. (2023) demostraron que la fosforilación jerárquica de *HOXB13* modula su estabilidad proteica y su capacidad de actuar como cofactor del AR, amplificando su actividad transcripcional oncogénica. A nivel de organización nuclear, Ahmed et al. (2021) evidenciaron que *HOXB13* se asocia a elementos reguladores de riesgo (*risk-associated cis-regulatory elements*, rCREs), destacando su papel central en la arquitectura tridimensional del genoma y en redes transcripcionales críticas para el cáncer de próstata.

Desde una perspectiva terapéutica, esta relevancia biológica contrasta con una limitación fundamental: *HOXB13* pertenece a la clase de factores de transcripción carentes de dominios catalíticos accesibles para la inhibición directa mediante pequeñas moléculas (Barashi et al., 2024; Cui et al., 2025). Estudios recientes han confirmado que *HOXB13* constituye una diana oncogénica farmacológicamente “indrogable”, altamente expresada en cáncer de próstata

metastásico resistente a la castración (Cui et al., 2025). En concordancia, Barashi et al. (2024) demostraron que la modulación de redes transcripcionales reguladas por *HOXB13* requiere enfoques indirectos, como la inhibición de proteínas epigenéticas asociadas, lo que subraya la necesidad de estrategias alternativas para interferir de forma directa con su función génica.

En conjunto, la evidencia disponible establece que *HOXB13* es un oncogén validado genética y funcionalmente en cáncer de próstata (Crawford et al., 2024; Ewing et al., 2012; Lu et al., 2022; Xiang et al., 2025). Sin embargo, su naturaleza como factor de transcripción no-drogable plantea una barrera significativa para la farmacología convencional, generando la necesidad de enfoques innovadores de silenciamiento génico (Crawford et al., 2024; Ewing et al., 2012; Lu et al., 2022; Xiang et al., 2025).

1.2. Introducción

Dada la relevancia de *HOXB13* como oncogén central en el cáncer de próstata y las limitaciones inherentes para su inhibición farmacológica directa, resulta necesario explorar estrategias alternativas capaces de modular su expresión génica de manera específica y controlada (Barashi et al., 2024; Cui et al., 2025). En este contexto, las tecnologías de silenciamiento génico emergen como una solución racional para superar la barrera terapéutica asociada a factores de transcripción no-drogables (Barashi et al., 2024; Cui et al., 2025).

Entre estas estrategias, la interferencia mediada por *Repeticiones Palindrómicas Cortas Agrupadas y Regularmente Interespaciadas* (CRISPR, *Clustered Regularly Interspaced Short Palindromic Repeats*), en su modalidad *CRISPR interferente* (CRISPRi, *CRISPR interference*), se ha consolidado como una plataforma robusta para la represión transcripcional dirigida (Ahmed et al., 2021; Dong & Kantor, 2021). (Dong & Kantor, 2021) describieron que CRISPRi permite la inhibición génica reversible mediante el reclutamiento de complejos que inducen heterocromatina, sin generar roturas de doble cadena en el ADN, lo que minimiza los riesgos asociados a la edición genómica convencional.

Desde el punto de vista técnico, CRISPRi se basa en el uso de una variante *Cas9 inactiva* (dCas9, *dead Cas9*) fusionada a dominios represores, como *Caja Asociada a Krüppel* (KRAB, *Krüppel-Associated Box*), guiada por un *Ácido Ribonucleico guía único* (sgRNA, *single-guide RNA*) hacia regiones regulatorias específicas del gen diana (Ahmed et al., 2021; Cai et al., 2023). Ahmed et al. (2021) y Cai et al. (2023) demostraron que este sistema permite una supresión transcripcional robusta y reversible mediante el reclutamiento de modificadores epigenéticos, sin alterar la secuencia del *Ácido Desoxirribonucleico* (DNA, *Deoxyribonucleic Acid*), lo que lo convierte en una herramienta adecuada para estudios funcionales y aplicaciones terapéuticas potenciales.

La aplicabilidad de CRISPRi en el contexto del cáncer de próstata ha sido validada en modelos celulares relevantes (Tian et al., 2023; Xiang et al., 2025; Yang et al., 2021). Yang et al. (2021) emplearon exitosamente el sistema dCas9–KRAB para la inhibición génica específica en la línea celular del cáncer de próstata (LNCaP), utilizada en estudios prostáticos. De forma similar, Tian et al. (2023) demostraron la utilidad de CRISPRi para interferir con elementos regulatorios asociados a loci de riesgo de cáncer de próstata, evidenciando su capacidad para modular regiones no codificantes funcionalmente relevantes. De manera directa, Xiang et al. (2025) mostraron que el silenciamiento de *HOXB13* mediante CRISPRi y siRNA reduce significativamente la expresión de AR y *Variante de Empalme 7 del Receptor de Andrógenos* (AR-V7), incluso en modelos resistentes, lo que refuerza su potencial como diana funcional de esta tecnología.

No obstante, la eficacia y especificidad de CRISPRi no son inherentes al sistema, sino que dependen del diseño racional de sus componentes moleculares, sgRNA (Javaid & Choi, 2021; Konstantakos et al., 2022; Listgarten et al., 2018). Konstantakos et al. (2022) señalaron que el desempeño de sistemas CRISPR-Cas depende de la secuencia y localización del sgRNA, mientras que Listgarten et al. (2018) identificaron la mitigación de efectos fuera de objetivo como uno de los principales cuellos de botella para la aplicación segura de estas tecnologías. En este sentido, (Javaid & Choi (2021) destacaron que la validación *in silico* de sgRNAs mediante predictores de eficiencia y especificidad constituye un paso crítico para garantizar el éxito de experimentos de interferencia génica.

En consecuencia, para sentar las bases de una futura herramienta de investigación orientada a evaluar la viabilidad de inhibir *HOXB13* como estrategia terapéutica en cáncer de

próstata, la presente tesis se propone realizar el diseño *in silico* de un sistema CRISPRi completo y específico dirigido contra este oncogén, abordando de manera sistemática los criterios de eficiencia, especificidad y seguridad que condicionan su aplicación.

1.3. Planteamiento del problema

El potencial de la inhibición transcripcional de *HOXB13* como estrategia para el estudio y eventual abordaje del cáncer de próstata ha sido respaldado por evidencia genética, funcional que lo posiciona como un oncogén central en la progresión tumoral, incluyendo contextos de enfermedad avanzada y resistente al tratamiento hormonal (Ewing et al., 2012; Lu et al., 2022; Xiang et al., 2025). Sin embargo, la traducción de este conocimiento biológico en una estrategia de silenciamiento génico efectiva requiere el desarrollo de un sistema CRISPRi preciso, específico y funcional, cuyo diseño no puede abordarse de manera empírica sin una evaluación computacional previa rigurosa (Javaid & Choi, 2021).

El diseño de un sistema CRISPRi dirigido contra *HOXB13* enfrenta desafíos fundamentales asociados a la complejidad regulatoria del locus génico (Ahmed et al., 2021; Dietlein et al., 2022). Diversos estudios han demostrado que la regulación transcripcional en cáncer de próstata no se limita a regiones promotoras proximales, sino que involucra elementos *cis*-regulatorios distales, como *enhancers* (*potenciador*), cuya actividad depende del contexto celular y del estado epigenético (Ahmed et al., 2021; Dietlein et al., 2022). En este escenario, un enfoque basado en la proximidad al *sitio de inicio de la transcripción* (TSS, *Transcription Start Site*) puede resultar subóptimo si no se consideran regiones reguladoras activas relevantes en cáncer de próstata (Ahmed et al., 2021; Dietlein et al., 2022).

Además, la eficacia del sistema CRISPRi depende de manera crítica del diseño del sgRNA, el cual determina tanto la eficiencia de silenciamiento dentro del objetivo (*on-target*) como el riesgo de efectos fuera de objetivo (*off-target*) (Konstantakos et al., 2022; Listgarten et al., 2018). La literatura ha demostrado que parámetros como la composición de secuencia, la estructura secundaria del sgRNA y el contexto genómico de la diana influyen de forma directa en la capacidad de unión y represión transcripcional mediada por dCas9-KRAB (Konstantakos et al., 2022; Riesenber et al., 2022). De forma paralela, la identificación y exclusión sistemática de sitios *off-target* constituye un requisito indispensable para evitar interpretaciones erróneas derivadas del silenciamiento de genes no deseados (Listgarten et al., 2018).

Finalmente, la implementación experimental de un sistema CRISPRi funcional requiere la integración de sus componentes moleculares —dCas9-KRAB y el sgRNA seleccionado— en un vector de expresión adecuado (Dong & Kantor, 2021; Yang et al., 2021). El diseño *in silico* de constructos plasmídicos que garanticen la expresión coordinada, estable y reproducible de estos elementos constituye una etapa crítica para asegurar la viabilidad experimental futura del sistema, reduciendo errores de clonación y fallos funcionales posteriores (Dong & Kantor, 2021; Yang et al., 2021).

En conjunto, el problema central de esta investigación radica en la ausencia de un diseño computacional integral, validado y reproducible de un sistema CRISPRi dirigido a *HOXB13*, que aborde la identificación de regiones regulatorias funcionales, el diseño optimizado de sgRNAs y la construcción racional de un vector expresable. Este vacío metodológico limita el desarrollo de estrategias de silenciamiento génico precisas contra *HOXB13* y justifica la necesidad de un enfoque *in silico* riguroso previo a cualquier implementación experimental.

1.4. Justificación

La elección de *HOXB13* como diana de estudio se encuentra respaldada por evidencia genética y funcional (Crawford et al., 2024; Ewing et al., 2012; Lu et al., 2022; Xiang et al., 2025). A nivel hereditario, variantes germinales como G84E confieren un incremento significativo del riesgo de cáncer de próstata, en casos de inicio temprano, posicionando a *HOXB13* como uno de los determinantes genéticos de mayor impacto clínico identificados hasta la fecha (Crawford et al., 2024; Ewing et al., 2012). Esta relevancia genética se ve reforzada por su inclusión en esquemas de evaluación clínica del riesgo hereditario (Dupont et al., 2021).

Desde una perspectiva funcional, *HOXB13* actúa como un nodo central en redes transcripcionales y metabólicas que sostienen la progresión tumoral prostática (Ahmed et al., 2021; Lu et al., 2022; Xiang et al., 2025). Esta función se enmarca en un comportamiento más amplio de la familia *HOX*, cuyos miembros desempeñan roles esenciales durante el desarrollo embrionario temprano, pero permanecen silenciados o con expresión mínima en la mayoría de los tejidos adultos diferenciados. Sin embargo, como documentaron Bhatlekar et al. (2014), múltiples genes *HOX* pueden reactivarse en contextos patológicos, particularmente en cáncer, donde participan en la regulación de proliferación celular, diferenciación y mantenimiento del estado tumoral. Este mismo trabajo demostró que los genes *HOX* no poseen una función

universal, sino que su actividad puede ser oncogénica o supresora dependiendo del contexto tisular y del estadio de la enfermedad, lo que explica por qué su reexpresión en tejidos adultos suele asociarse a procesos neoplásicos más que a funciones fisiológicas normales.

En el caso específico de *HOXB13*, su expresión elevada en tejidos prostáticos adultos no refleja un programa de desarrollo normal, sino la activación aberrante de redes transcripcionales propias del cáncer de próstata. Estudios experimentales han demostrado que *HOXB13* regula tanto programas dependientes del receptor de andrógenos como mecanismos independientes del AR, incluyendo el control epigenético de la lipogénesis *de novo*, procesos directamente asociados a la agresividad tumoral y resistencia terapéutica (Lu et al., 2022; Xiang et al., 2025). En consecuencia, la inhibición transcripcional de *HOXB13* no implica interferir con un regulador esencial del desarrollo en un contexto fisiológico, sino modular un factor reactivado de forma patológica en estadios adultos, lo que justifica su abordaje como diana terapéutica y experimental en cáncer de próstata (Lu et al., 2022; Xiang et al., 2025).

No obstante, *HOXB13* presenta una limitación terapéutica fundamental: su naturaleza como factor de transcripción carente de dominios catalíticos accesibles lo clasifica como una diana farmacológicamente “no-drogable” (Barashi et al., 2024; Cui et al., 2025). Esta característica ha sido documentada en cáncer de próstata metastásico resistente a la castración, donde *HOXB13* tiene una alta expresión y contribuye a la progresión de la enfermedad (Cui et al., 2025). Por ello, resulta imperativo explorar plataformas alternativas capaces de modular su expresión génica sin requerir inhibición proteica directa (Cui et al., 2025; Dong & Kantor, 2021).

La interferencia mediada por CRISPR (CRISPRi) constituye una plataforma adecuada para abordar genes no-drogables, al permitir la represión transcripcional específica sin inducir roturas de doble cadena en el ADN, reduciendo riesgos asociados a mutagénesis genómica (Ahmed et al., 2021; Dong & Kantor, 2021). Sin embargo, la literatura coincide en que el éxito de cualquier aplicación CRISPR depende de una etapa previa de diseño *in silico*, orientada a maximizar eficiencia, especificidad y seguridad (Javaid & Choi, 2021; Konstantakos et al., 2022).

En este sentido, el uso combinado de herramientas bioinformáticas consolidadas responde a las mejores prácticas actuales del campo (Hoberecht et al., 2022; Javaid & Choi,

2021). Plataformas como CHOPCHOP y CRISPOR integran algoritmos validados para la predicción de eficiencia *on-target* y evaluación inicial de especificidad de sgRNAs (Konstantakos et al., 2022; Tian et al., 2023; Yang et al., 2021). De manera complementaria, Cas-OFFinder permite la identificación exhaustiva de sitios *off-target* potenciales con distintos niveles de desajuste, superando enfoques basados en alineamientos básicos (Cai et al., 2023; Yi et al., 2022). Asimismo, el análisis de estructura secundaria mediante RNAfold resulta esencial para descartar sgRNAs con conformaciones estructurales desfavorables que comprometan su estabilidad y eficacia funcional (Javaid & Choi, 2021; Riesenber et al., 2022).

La identificación de secuencias reguladoras funcionales en el locus de *HOXB13* se justifica por la necesidad de considerar regiones de cromatina activa, más allá del TSS, que reflejen el contexto epigenético real del cáncer de próstata y maximicen la probabilidad de un silenciamiento efectivo mediante CRISPRi (Ahmed et al., 2021; Dietlein et al., 2022). El diseño y análisis exhaustivo de sgRNAs se fundamenta en la necesidad de equilibrar eficacia y seguridad, dado que un diseño subóptimo puede conducir a resultados experimentales erróneos o a artefactos derivados de efectos *off-target* (Listgarten et al., 2018). Finalmente, la construcción *in silico* de un vector de expresión funcional se justifica por la necesidad de asegurar la traducción práctica del diseño teórico a un sistema genético utilizable en futuros estudios experimentales (Dong & Kantor, 2021; Yang et al., 2021).

En conjunto, esta investigación se justifica porque aborda una diana biológica y clínicamente relevante mediante una tecnología de vanguardia cuyo éxito depende de una etapa de diseño computacional rigurosa. El pipeline *in silico* propuesto constituye un paso fundacional e indispensable para futuros estudios experimentales orientados a evaluar el potencial de *HOXB13* como diana terapéutica en cáncer de próstata.

1.5. Objetivos

1.5.1. *Objetivo general*

Diseñar *in silico* un sistema CRISPRi dirigido a inhibir la expresión del oncogén *HOXB13* en cáncer de próstata.

1.5.2. *Objetivos específicos*

1. Identificar las secuencias reguladoras del gen *HOXB13* como posibles dianas para la inhibición mediante CRISPRi.
2. Analizar *in silico* guías de ARN (sgRNAs) específicas y eficientes que dirijan el complejo dCas9-KRAB hacia el oncogén *HOXB13*.
3. Construir el sistema CRISPRi en un vector plasmídico que integre dCas9-KRAB y la sgRNA seleccionada.

1.6. Hipótesis

El diseño *in silico* de un sistema CRISPRi basado en dCas9-KRAB permitirá proponer un modelo computacional de silenciamiento transcripcional específico y eficiente del oncogén *HOXB13* en células prostáticas, constituyendo una estrategia potencial para modular su sobreexpresión en el cáncer de próstata.

CAPÍTULO 2

Marco teórico

2.1. Los genes *HOX* como reguladores del desarrollo y su implicación en cáncer

Los genes homeobox (*HOX*) constituyen una familia de factores de transcripción caracterizados por la presencia de un dominio de unión al DNA conservado, el homeodominio, cuya función principal es la regulación de programas transcripcionales durante el desarrollo embrionario (Bhatlekar et al., 2014). En humanos, estos genes se organizan en cuatro clústeres (*HOXA*, *HOXB*, *HOXC* y *HOXD*) distribuidos en distintos cromosomas, manteniendo un principio de colinealidad entre su posición genómica y su patrón de expresión espacio-temporal (Bhatlekar et al., 2014).

Bhatlekar et al. (2014) documentan que múltiples miembros de la familia *HOX* permanecen activos o son reactivados en tejidos adultos bajo condiciones patológicas, incluyendo diversos tipos de cáncer, donde participan en la regulación de proliferación celular, diferenciación y mantenimiento del estado tumoral. El mismo trabajo demuestra que los genes *HOX* pueden desempeñar roles oncogénicos o supresores tumorales dependiendo del contexto tisular específico, estableciendo que su función no es universal sino altamente dependiente del entorno celular (Bhatlekar et al., 2014).

Dentro de esta organización genómica, *HOXB13* pertenece al clúster *HOXB*, localizado en el cromosoma 17, y corresponde a uno de los genes terminales del clúster, una característica asociada a funciones especializadas en tejidos diferenciados más que a patrones embrionarios tempranos (Bhatlekar et al., 2014).

2.2. *HOXB13* como regulador crítico en células de cáncer de próstata

El papel de *HOXB13* en cáncer de próstata está respaldado por evidencia genética directa. Shang et al. (2013), mediante un meta-análisis que incluyó más de 24 000 casos y 73

000 controles, demostraron que la mutación germinal G84E en *HOXB13* se asocia de forma significativa con un aumento del riesgo de cáncer de próstata en poblaciones de ascendencia europea, con odds ratios elevados y estadísticamente robustos. Este estudio demuestra que *HOXB13* no es un marcador tisular, sino un gen cuya alteración heredada contribuye directamente a la susceptibilidad a la enfermedad (Shang et al., 2013).

A nivel funcional, Lu et al. (2022) demostraron experimentalmente que *HOXB13* regula el metabolismo lipídico en células de cáncer de próstata mediante la supresión de la lipogénesis *de novo*. Este trabajo evidencia que *HOXB13* recluta a HDAC3 para inducir un estado epigenético represivo en genes clave de la vía lipogénica, y que la inhibición de *HOXB13* revierte este programa metabólico, alterando la viabilidad y el comportamiento tumoral de las células prostáticas (Lu et al., 2022). Estos resultados establecen un rol oncogénico activo y definido de *HOXB13* en cáncer de próstata.

El carácter central de *HOXB13* dentro de redes transcripcionales prostáticas también ha sido demostrado a nivel de interacciones proteína–proteína. Erokhin et al. (2025) identificaron el interactoma de *HOXB13* en células de cáncer de próstata mediante purificación por inmunoafinidad seguida de espectrometría de masas, demostrando que *HOXB13* interactúa con múltiples factores de transcripción, incluyendo *TBX3*. Este estudio mostró que la pérdida funcional de *HOXB13* y *TBX3* afecta de manera crítica la proliferación y supervivencia de las mismas líneas celulares de cáncer de próstata, evidenciando una dependencia funcional compartida (Erokhin et al., 2025).

El carácter central de *HOXB13* en estadios avanzados de la enfermedad fue demostrado por Z. Chen et al. (2018), quienes evidenciaron que *HOXB13* gobierna el cistroma del receptor androgénico variante AR-V7 en cáncer de próstata resistente a la castración. Este trabajo mostró que la ocupación genómica de AR-V7 depende de *HOXB13* y que la pérdida de este factor altera programas transcripcionales característicos del cáncer de próstata resistente, demostrando su actividad funcional en contextos independientes de andrógenos (Chen et al., 2018).

Adicionalmente, Erokhin et al. (2025) caracterizaron el interactoma de *HOXB13* en células de cáncer de próstata mediante purificación por inmunoafinidad seguida de espectrometría de masas, demostrando que *HOXB13* interactúa con múltiples factores de transcripción, incluyendo *TBX3*. Este estudio mostró que la pérdida funcional de *HOXB13* y

TBX3 afecta de manera crítica la proliferación y supervivencia de las mismas líneas celulares de cáncer de próstata, evidenciando una dependencia funcional compartida (Erokhin et al., 2025).

2.3. *HOXB13* como diana transcripcional y limitaciones farmacológicas

Los estudios experimentales disponibles demuestran que *HOXB13* actúa como factor de transcripción, ejerciendo su función oncogénica mediante la regulación directa de programas transcripcionales y epigenéticos, así como a través de interacciones proteína-ADN y proteína-proteína (Chen et al., 2018; Erokhin et al., 2025; Lu et al., 2022). Esta modalidad de acción implica la ausencia de una actividad enzimática intrínseca susceptible de inhibición directa mediante fármacos de molécula pequeña.

En consecuencia, la evidencia disponible indica que estrategias dirigidas a la modulación de la actividad de *HOXB13* deben operar a nivel génico o transcripcional, más que a nivel proteico, para lograr una interferencia funcional efectiva de su papel en la progresión tumoral.

2.4. Interferencia CRISPR como estrategia de silenciamiento transcripcional

Qi et al. (2013) demostraron que dCas9, guiada por un ARN específico, puede unirse a secuencias génicas definidas y bloquear la transcripción sin inducir roturas en el ADN, estableciendo el concepto de CRISPR. Este estudio mostró que el silenciamiento es específico, dependiente de la secuencia del sgRNA y reversible (Qi et al., 2013).

Posteriormente, Gilbert et al. (2013) demostraron que la fusión de dCas9 a dominios represores como KRAB incrementa de manera significativa la eficacia del silenciamiento transcripcional en células eucariotas, permitiendo una represión estable y específica de genes endógenos sin generar modificaciones permanentes en la secuencia genómica. Este trabajo validó CRISPRi como una plataforma modular de regulación génica en células humanas (Gilbert et al., 2013a).

La aplicabilidad de CRISPRi en sistemas complejos fue demostrada mediante su uso para la inactivación específica de genes en tejidos *in vivo*, evidenciando que el sistema puede operar de forma eficaz fuera de modelos celulares simples (Zheng et al., 2018).

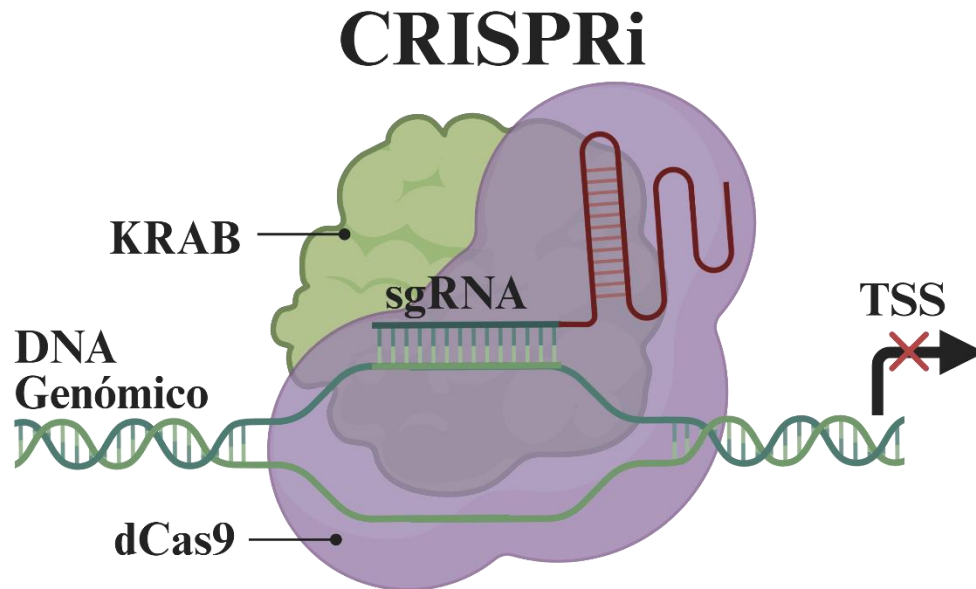


Figura 1. Esquema del sistema CRISPRi. Elaborado en Biorender (Created in <https://BioRender.com>).

2.5. Justificación del diseño *in silico* del sistema CRISPRi contra *HOXB13*

La evidencia experimental disponible demuestra que *HOXB13* es un regulador transcripcional activo en cáncer de próstata, cuya alteración genética incrementa el riesgo de enfermedad (Shang et al., 2013), cuya actividad epigenética y metabólica sostiene el fenotipo tumoral (Lu et al., 2022), y cuya función transcripcional es crítica en estadios resistentes a la castración (Chen et al., 2018), además de su integración en redes transcripcionales esenciales para la supervivencia celular (Erokhin et al., 2025). De manera paralela, CRISPRi constituye una herramienta demostrada para el silenciamiento específico, reversible y eficiente de genes a nivel transcripcional en células humanas (Gilbert et al., 2013a; Qi et al., 2013).

Sin embargo, la eficacia del sistema CRISPRi ha demostrado depender de manera significativa de las características del sgRNA, incluyendo su localización genómica y su secuencia, las cuales influyen en la magnitud del silenciamiento transcripcional observado

(Gilbert et al., 2013a; Qi et al., 2013). En estos estudios, se evidencia que diferentes sgRNAs dirigidos a un mismo gen pueden producir niveles variables de represión, subrayando la importancia de una selección cuidadosa de las guías.

En este contexto, el diseño in silico de sgRNAs dirigidos contra *HOXB13* constituye una etapa fundacional para maximizar la eficacia del silenciamiento transcripcional mediante CRISPRi. Por lo tanto, este marco teórico sustenta el objetivo de la presente tesis: diseñar y validar in silico un sistema CRISPRi dirigido al silenciamiento específico de *HOXB13*, como paso previo indispensable para futuras aplicaciones experimentales y terapéuticas en cáncer de próstata resistente.

CAPÍTULO 3

Metodología

3.1. Fase 1: Caracterización genómica del gen

3.1.1. Identificación del gen HOXB13 y extracción de la región diana en UCSC Genome Browser

La Fase 1 consistió en la identificación y documentación del locus genómico del gen *HOXB13* en *Homo sapiens*, así como en la extracción de una región específica del genoma para su uso en las etapas posteriores del diseño *in silico* de sgRNAs para CRISPRi. Para este propósito se utilizó el UCSC Genome Browser (James Kent et al., 2002) , empleando el ensamblaje genómico Human Dec. 2013 (GRCh38/hg38), con el fin de mantener consistencia con las herramientas bioinformáticas utilizadas en las fases siguientes (Casper et al., 2026; James Kent et al., 2002).

La identificación del gen *HOXB13* se realizó a partir del transcrito de referencia RefSeq NM_006361.6, clasificado como proteína codificante (*protein_coding*, registro revisado). A partir de esta anotación se registraron los parámetros básicos del locus, incluyendo cromosoma, banda citogenética, coordenadas genómicas y orientación de hebra. El gen *HOXB13* se localiza en el cromosoma 17 (17q21.32) y se encuentra anotado en orientación negativa (-).

Con base en la anotación del transcrito de referencia, se delimitó una región genómica proximal al sitio de inicio de la transcripción, correspondiente al intervalo chr17:48728450–48728800 (hg38). La secuencia fue extraída mediante la herramienta “Get DNA in Window” del UCSC Genome Browser y exportada en formato FASTA, sin aplicar enmascaramiento de repeticiones ni adicionar bases en los extremos 5’ o 3’. Los metadatos asociados a la extracción fueron registrados para asegurar la trazabilidad y reproducibilidad del flujo de trabajo *in silico*. Esta región constituyó la secuencia de entrada para las herramientas de diseño de sgRNAs empleadas en la Fase 2.

Tabla 1. Parámetros de caracterización genómica del gen *HOXB13* y definición de la región diana.

Parámetro	Descripción
Organismo	<i>Homo sapiens</i>
Ensamblaje genómico	Human Dec. 2013 (GRCh38/hg38)
Gen	<i>HOXB13</i>
Transcrito de referencia	RefSeq NM_006361.6
Tipo de transcrito	Proteína codificante (protein_coding)
Estado del registro	Revisado (Reviewed)
Cromosoma	17
Banda citogenética	17q21.32
Orientación de hebra	Negativa (-)
Coordenadas del transcrito	chr17:48724763–48728750
Región extraída para análisis	chr17:48728450–48728800
Tamaño de la región extraída	351 pb
Herramienta de extracción	<i>Get DNA in Window</i>
Formato de salida	FASTA
Enmascaramiento de repeticiones	No aplicado
Bases adicionales 5' / 3'	0 / 0

3.2. Fase 2: Diseño y preselección bioinformática de ARN guía (sgRNAs) dirigidos a *HOXB13*

3.2.1. Generación y preselección de crRNAs candidatos mediante CHOPCHOP

El diseño inicial de las secuencias guía dirigidas al gen *HOXB13* se realizó utilizando la plataforma web CHOPCHOP (Labun et al., 2025), como parte de la Fase 2 del flujo metodológico *in silico*. CHOPCHOP permite la identificación y ARNs CRISPR (crRNA, *CRISPR RNA*) compatibles con sistemas CRISPR/Cas, integrando métricas de diseño y un análisis preliminar de coincidencias genómicas potenciales para apoyar la selección inicial de candidatos.

Como secuencia de entrada se utilizó la región genómica definida en la Fase 1, correspondiente a la región regulatoria seleccionada del gen *HOXB13* en el genoma humano. La búsqueda se configuró para *Homo sapiens* (*GRCh38/hg38*), seleccionando el sistema *CRISPRi*, con longitud de guía de 20 nucleótidos y motivo *PAM 5'-NGG-3'*, compatible con SpCas9/dCas9. Los parámetros generales empleados en CHOPCHOP se resumen en la Tabla 2.

Tabla 2. Configuración general empleada en CHOPCHOP para el diseño de crRNAs dirigidas a *HOXB13*.

Objetivo	En	Usando	Para	Longitud del crRNA sin PAM	PAM-3'
Resultado fase1	<i>Homo sapiens</i> (<i>hg38/GRCh38</i>)	CRISPR/Cas9	Represión	20nt	NGG

A partir del conjunto total de crRNAs generados por CHOPCHOP, se realizó una preselección inicial de candidatos, aplicando criterios coherentes con un enfoque *CRISPRi*. En primer lugar, se priorizó la localización dentro de la ventana regulatoria definida en la Fase 1, considerando que la represión transcripcional mediada por dCas9-KRAB es más eficiente cuando el complejo se posiciona en regiones próximas al TSS.

En segundo lugar, se exigió una longitud exacta de 20 nt, la presencia del PAM NGG y un contenido de GC dentro de un rango operativo aproximado (40–70%), evitando extremos

que puedan comprometer la estabilidad del apareamiento o favorecer plegamientos no deseados (Wang et al., 2014). Adicionalmente, se priorizaron crRNAs con baja autocomplementariedad con el fin de reducir la formación de estructuras intramoleculares que puedan interferir con la funcionalidad del complejo CRISPR.

La evaluación comparativa se ejecutó bajo un esquema jerárquico de decisión, manteniendo la selección final abierta para validación posterior mediante herramientas específicas de análisis off-target. El orden de prioridad aplicado fue: (i) **Proximidad al TSS**, priorizando crRNAs localizados dentro de una ventana estrecha alrededor del sitio de inicio transcripcional (± 50 pb), región considerada óptima para CRISPRi (Gilbert et al., 2014); (ii) **Perfil preliminar de off-targets**, penalizando de manera estricta aquellas secuencias con coincidencias genómicas que presentaran entre 0 y 2 desajustes, especialmente cuando estos conservaban complementariedad en la región PAM-proximal (Fu et al., 2013); (iii) **Autocomplementariedad**, priorizando crRNAs con emparejamientos internos mínimos, de acuerdo con evidencia experimental que demuestra que estructuras secundarias estables reducen la eficiencia del complejo Cas9-gRNA (Thyme et al., 2016; Konstantakos et al., 2022); y (iv) **Eficiencia predictiva**, empleada como criterio de desempate utilizando modelos derivados de Doench et al. (2014).

Esta priorización se justificó debido a que los efectos off-target representan un riesgo metodológico relevante en aplicaciones CRISPR, y se ha demostrado que enfoques *in silico* basados en conteos de desajustes (*mismatches*) pueden subestimar sitios potenciales si no consideran variantes de alineamiento más complejas (Labun et al., 2025).

Finalmente, el conjunto de crRNAs preseleccionados mediante CHOPCHOP fue sometido a una evaluación más estricta de especificidad y estabilidad estructural utilizando herramientas complementarias CRISPOR (Concordet & Haeussler, 2018), Cas-OFFinder (Bae et al., 2014) y RNAfold (Lorenz et al., 2011), con el objetivo de sustentar una selección final basada en criterios integrados de eficiencia, especificidad y seguridad genómica antes de su implementación en un sistema CRISPRi completo.

En resumen, el conjunto de crRNAs preseleccionados mediante CHOPCHOP fue sometido a una evaluación más estricta de especificidad y estabilidad estructural utilizando herramientas complementarias (CRISPOR (Concordet & Haeussler, 2018), Cas-OFFinder (Bae

et al., 2014) y RNAfold (Lorenz et al., 2011), con el objetivo de sustentar una selección final basada en criterios integrados de eficiencia, especificidad y seguridad genómica antes de su implementación en un sistema CRISPRi completo.

3.2.2. Determinación de coordenadas genómicas y cálculo de distancia al TSS para crRNAs candidatas

La localización genómica precisa de cada crRNA candidata y su distancia relativa al TSS del gen *HOXB13* se establecieron como criterios cuantitativos centrales para la selección de guías en el contexto de CRISPRi, dado que la eficacia de la interferencia transcripcional mediada por dCas9-KRAB depende críticamente del posicionamiento respecto al TSS. Este procedimiento se desarrolló siguiendo y extendiendo los principios de anotación y posicionamiento genómico descritos por Haeussler et al. (2016).

Como referencia se utilizaron los parámetros definidos en la Fase 1 (Sección 3.1.1 y Tabla 1), incluyendo las coordenadas de la región genómica extraída (chr17:48 728 450–48 728 800, ensamblaje GRCh38/hg38) y la posición del TSS del transcrito de referencia NM_006361.6 de *HOXB13*, localizado en la coordenada chr17:48 728 750. Dado que *HOXB13* se transcribe desde la hebra negativa, esta coordenada corresponde al extremo 5' funcional del gen.

CHOPCHOP reporta la ubicación de cada crRNA mediante una posición relativa dentro de la secuencia analizada (notación seq:XXX), correspondiente al primer nucleótido de la guía dentro del fragmento de 351 pb. Para convertir estas posiciones relativas en coordenadas genómicas absolutas compatibles con UCSC Genome Browser y CRISPOR, se aplicó la siguiente relación:

$$\text{Coordenada genómica} = \text{Inicio de la región extraída} + (\text{posición seq} - 1)$$

Donde: el inicio de la región extraída corresponde a la coordenada 48 728 450. Este ajuste permite mantener coherencia con el sistema de coordenadas basado en 1 utilizado por UCSC Genome Browser.

La distancia de cada crRNA respecto al TSS se calculó como:

$$\text{Distancia al TSS} = \text{Coordenada genómica} - \text{TSS}$$

En esta convención, valores negativos indican localización aguas arriba del TSS (región promotora), valores positivos indican localización aguas abajo (región transcrita), y el valor cero corresponde a coincidencia exacta con el nucleótido del TSS.

Para la priorización de crRNAs en el diseño CRISPRi, se consideraron preferentes aquellas localizadas dentro de una ventana de -200 a $+50$ pb respecto al TSS, con énfasis en la región promotora proximal (-50 a $+20$ pb), reportada como óptima para maximizar la represión transcripcional mediada por dCas9-KRAB. Cuando dos o más crRNAs presentaron perfiles de especificidad genómica comparables, se otorgó prioridad adicional a los localizados en la hebra no codificante (-), en concordancia con el sentido transcripcional del gen.

Este procedimiento permitió integrar de forma reproducible y cuantitativa la posición relativa al TSS como un criterio estructural de selección, complementando los análisis de especificidad genómica y estabilidad estructural desarrollados en las etapas posteriores del diseño *in silico*.

3.2.3. Evaluación de especificidad genómica y depuración avanzada de crRNAs mediante CRISPOR

Las secuencias crRNA preseleccionadas en la etapa previa mediante CHOPCHOP fueron evaluadas utilizando la plataforma CRISPOR (Concordet & Haeussler, 2018), con el objetivo de estimar su especificidad genómica y caracterizar el riesgo de efectos off-targets. Esta etapa se consideró crítica debido a que, en sistemas CRISPRi, la unión inespecífica del complejo dCas9-KRAB puede inducir represión transcripcional no deseada en loci no diana.

Para cada crRNA, la secuencia de 20 nucleótidos fue ingresada una a una en el Step 1 de CRISPOR. En el Step 2 se seleccionó el genoma de referencia *Homo sapiens* (GRCh38/hg38; Human Dec. 2013). En el Step 3 se configuró el sistema *SpCas9* (20 pb-NGG), empleando la salida estándar del servidor, la cual reporta puntajes de especificidad, predicción

de actividad on-target y un listado de sitios off-target potenciales en el genoma. El análisis se realizó guía por guía para conservar trazabilidad y permitir la comparación directa entre candidatas.

Las crRNAs que presentaron advertencias de incompatibilidad con el genoma de referencia o que no pudieron ser mapeadas de forma confiable por la plataforma fueron excluidas del análisis posterior, al no cumplir con los requisitos de validación genómica bajo el ensamblaje hg38.

La interpretación de los resultados se efectuó mediante un esquema de criterios jerarquizados. Como criterio primario se evaluó la especificidad, utilizando el *Puntaje del Instituto Tecnológico de Massachusetts* (MIT, *Massachusetts Institute of Technology Score*) y el *Puntaje de Determinación de la Frecuencia de Corte* (CFD, *Cutting Frequency Determination Score*) reportados por CRISPOR como métricas operativas para priorizar crRNAs con menor riesgo off-target (Concordet & Haeussler, 2018; Listgarten et al., 2018). Se establecieron como umbrales mínimos aceptables valores de MIT ≥ 90 % y CFD ≥ 94 –96 %. También, se consideró el número total de sitios off-target detectados, priorizando candidatas sin coincidencias críticas y excluyendo aquellas con sitios potenciales que presentaran ≤ 1 desajuste respecto a la secuencia diana, en particular cuando dichos desajustes se localizaban en la región PAM-proximal (Fu et al., 2013).

Cuando dos crRNAs presentaron perfiles de especificidad comparables, se utilizó como criterio secundario su proximidad al TSS (derivada de la Fase 1), considerando una ventana aproximada de ± 50 pb, en coherencia con el diseño orientado a interferencia transcripcional mediante CRISPRi (Gilbert et al., 2014). La eficiencia predictiva on-target reportada por CRISPOR (modelo Doench 2016) se empleó como criterio de desempate, dado que CRISPRi no depende de la generación de roturas en el ADN (Doench et al., 2014).

Finalmente, las crRNAs que cumplieron con los criterios establecidos fueron retenidas para los análisis posteriores de depuración exhaustiva de off-targets mediante Cas-OFFinder (Bae et al., 2014) y de consistencia estructural con RNAfold, (Lorenz et al., 2011) previos a la priorización final.

3.2.4. Depuración exhaustiva de off-targets genómicos mediante Cas-OFFinder

Para la evaluación exhaustiva de sitios fuera de objetivo (off-targets), el análisis se basó en el trabajo fundacional de Bae et al. (2014), quienes desarrollaron Cas-OFFinder como un algoritmo flexible capaz de identificar coincidencias genómicas con múltiples desajustes sin las limitaciones de búsquedas convencionales. Asimismo, se tomaron como referencia los estudios de Cai et al. (2023) y Yi et al. (2022), que emplean Cas-OFFinder como herramienta central en flujos *in silico* de validación de especificidad para sistemas CRISPR, destacando su utilidad para la depuración conservadora de crRNAs antes de la implementación experimental. En conjunto, estos trabajos sustentan el uso de Cas-OFFinder en esta etapa metodológica para minimizar el riesgo de efectos fuera de objetivo en el diseño de sistemas CRISPRi dirigidos a genes humanos.

Cas-OFFinder se realizó la búsqueda de sitios potenciales de unión del complejo Cas9–crRNA sin restricciones estrictas en el número de desajustes y considerando la degeneración del motivo PAM. También, la plataforma fue utilizada para evaluar la especificidad de las crRNAs candidatas en el genoma humano de referencia *Homo sapiens* (GRCh38/hg38).

El análisis se configuró para el sistema SpCas9 derivado de *Streptococcus pyogenes*, especificando el motivo PAM 5'-NGG-3', compatible con aplicaciones CRISPR/Cas9 y CRISPRi en células humanas. Las secuencias de entrada correspondieron a las regiones guía de 20 nucleótidos, sin incluir la secuencia PAM, las cuales fueron introducidas una a una en el formato requerido por la plataforma.

Con el objetivo de identificar posibles sitios off-target de baja y muy baja similitud, el número máximo de desajustes permitidos se estableció en cuatro (4) mismatches, sin permitir inserciones ni deleciones, configurando los parámetros de DNA bulge y RNA bulge en valor cero. Esta configuración permitió un análisis estricto basado en sustituciones nucleotídicas, alineado con enfoques conservadores orientados a minimizar el riesgo de unión inespecífica del complejo dCas9–crRNA.

Los resultados generados por Cas-OFFinder incluyeron la localización cromosómica de cada coincidencia potencial, el número de desajustes respecto a la secuencia diana y la orientación de la unión. Para la interpretación de estos resultados se aplicó el siguiente criterio

de especificidad: una crRNA fue considerada específica cuando presentó un único sitio con 0 mismatches correspondiente al locus diana esperado (región regulatoria de *HOXB13* en el cromosoma 17) y no mostró coincidencias adicionales con 0, 1 o 2 desajustes en el resto del genoma.

Los sitios identificados con 3 o 4 mismatches fueron considerados de bajo impacto funcional, dado que la probabilidad de unión estable del complejo Cas9 disminuye con el incremento del número de desajustes, fuera de la región PAM-proximal. No obstante, estos sitios fueron documentados como parte del análisis integral de especificidad, manteniendo un enfoque conservador orientado a la bioseguridad del diseño.

Este análisis permitió refinar la selección de crRNAs candidatas, priorizando aquellas con el perfil más restrictivo de especificidad genómica antes de proceder a la evaluación de consistencia estructural mediante predicción de estructura secundaria en la etapa posterior del flujo metodológico.

3.2.5. Evaluación de la consistencia estructural del sgRNA mediante RNAfold

La consistencia estructural y el plegamiento secundario del sgRNA completo se evaluaron utilizando el servidor RNAfold (Gruber et al., 2008; Lorenz et al., 2011), herramienta basada en la predicción de estructuras secundarias de ARN mediante energía mínima libre (MFE, *Minimum Free Energy*) y el cálculo del ensamble termodinámico a través de la función de partición. Este enfoque se emplea de forma habitual en el diseño *in silico* de guías CRISPR para descartar configuraciones estructurales que puedan afectar la accesibilidad del ARN guía o la estabilidad del complejo dCas9-ARN.

El análisis estructural se incorporó debido a que la organización conformacional del ARN guía, en la región crRNA, puede influir en su accesibilidad para la hibridación con la secuencia diana de ADN, un aspecto relevante en aplicaciones de regulación transcripcional como CRISPRi (Doench et al., 2016).

Para cada candidato retenido tras los filtros de especificidad genómica (Ranks 1, 3, 6 y 10), se construyó una secuencia de entrada correspondiente al sgRNA completo, compuesta por la región crRNA de 20 nucleótidos y el scaffold estándar de sgRNA para sistemas

SpCas9/dCas9 descrito por (Cong et al., 2013). Las secuencias se ingresaron en orientación 5'→3' para garantizar un contexto estructural comparable entre candidatas.

En RNAfold se mantuvieron los parámetros termodinámicos por defecto y se activó el cálculo de la estructura de mínima energía libre (MFE) y del ensamble termodinámico completo mediante partition function. Adicionalmente, se seleccionaron las salidas gráficas de estructura secundaria interactiva y mountain plot para documentar el plegamiento y facilitar la comparación cualitativa entre guías.

A partir de la salida estándar del servidor se registraron cuatro métricas cuantitativas para cada sgRNA: (i) energía mínima libre de la estructura MFE (kcal/mol), (ii) energía libre del ensamble termodinámico (kcal/mol), (iii) frecuencia de la estructura MFE dentro del ensamble (%) y (iv) diversidad del ensamble (*ensemble diversity*). Estas variables se emplearon como criterios operativos de consistencia estructural.

Para la interpretación comparativa se definieron umbrales internos del estudio. La MFE se utilizó como indicador de estabilidad global, considerando favorables valores negativos dentro de un rango estrecho entre candidatas (-23 a -20 kcal/mol). La energía libre del ensamble se empleó como medida complementaria de estabilidad promedio, priorizando guías con diferencias moderadas entre ambos valores.

La diversidad del ensamble se interpretó como indicador de dispersión conformacional (baja: 10–16; media: 16–22; alta: >22), mientras que la frecuencia de la estructura MFE se utilizó como medida de representatividad del plegamiento dominante. Los gráficos de montaña (*mountain plots*) se conservaron como evidencia visual complementaria del grado de concordancia estructural entre las guías analizadas.

3.2.6. Integración de resultados de la Fase 2

Una vez completadas las evaluaciones individuales de las crRNAs mediante CHOPCHOP, CRISPOR, Cas-OFFinder y RNAfold, se realizó una integración sistemática de los resultados de la Fase 2 con el objetivo de consolidar en un único marco analítico los parámetros obtenidos para cada candidata.

Esta integración permitió unificar, por crRNA, métricas posicionales relativas al TSS, parámetros de especificidad genómica, perfiles exhaustivos de off-target y variables termodinámicas asociadas a la estabilidad estructural del sgRNA completo. Los datos integrados correspondieron exclusivamente a resultados *in silico* derivados de herramientas bioinformáticas deterministas, sin incluir inferencias experimentales.

La matriz resultante fue preparada como insumo para análisis comparativos posteriores, permitiendo la evaluación conjunta y coherente de los candidatos bajo criterios homogéneos. Esta etapa metodológica constituyó la base para la síntesis de resultados y para los análisis estadísticos exploratorios desarrollados en las secciones siguientes.

3.2.7. Análisis estadístico exploratorio y priorización multicriterio de crRNAs

3.2.7.1. Enfoque general del análisis estadístico

El análisis estadístico del presente estudio se realizó con el objetivo de comparar, explorar y priorizar crRNAs candidatas a partir de métricas bioinformáticas genómicas, estructurales y de especificidad obtenidas en la Fase 2. Dado que los datos corresponden a resultados *in silico*, generados por algoritmos deterministas y sin réplicas biológicas, se adoptó un enfoque descriptivo y no paramétrico, evitando el uso de métodos inferenciales que asumen normalidad o independencia experimental (Doench et al., 2016; Labun et al., 2019).

Todos los análisis estadísticos y las visualizaciones gráficas se realizaron utilizando el software R (RStudio Team, 2026), empleando principalmente funciones base del lenguaje. Esta elección permitió asegurar reproducibilidad, estabilidad del entorno de análisis y compatibilidad entre versiones, en coherencia con buenas prácticas de análisis computacional. Las crRNAs se trataron como unidades analíticas discretas, y los resultados estadísticos se reportaron de forma descriptiva y comparativa, en concordancia con el carácter exploratorio y *in silico* del estudio.

Este enfoque permitió evaluar tendencias, relaciones y patrones relativos entre crRNAs sin extrapolar conclusiones de tipo experimental o poblacional, manteniendo el alcance metodológico del trabajo dentro de los límites propios de un diseño computacional.

3.2.7.2. Preparación y estructuración de los datos

Los resultados generados por CHOPCHOP, CRISPOR, Cas-OFFinder y RNAfold fueron consolidados en una matriz única organizada por crRNA candidata. Las variables incluidas correspondieron a parámetros cuantitativos tales como: puntajes de especificidad MIT y CFD, número total de off-targets, energía mínima libre (MFE), frecuencia relativa de la estructura MFE, diversidad del ensamble conformacional, contenido GC y distancia al sitio de inicio de transcripción (TSS).

La organización y verificación de consistencia de los datos se realizó en el entorno estadístico R, comprobando la ausencia de valores faltantes y la coherencia numérica entre plataformas. Las crRNAs fueron tratadas como identificadores categóricos para los análisis comparativos y como etiquetas en las representaciones gráficas, mientras que las métricas asociadas se consideraron variables numéricas continuas.

3.2.7.3. Análisis comparativo no paramétrico

Para evaluar diferencias globales entre crRNAs respecto a los distintos parámetros genómicos y estructurales, se utilizó la prueba no paramétrica de Kruskal–Wallis, adecuada para tamaños muestrales pequeños y distribuciones no normales. Esta prueba se aplicó a variables como energía mínima libre (MFE), puntajes MIT y CFD, número total de off-targets y distancia al TSS.

El uso de esta prueba permitió determinar si las variaciones observadas entre crRNAs superaban lo esperado por variabilidad interna del conjunto, sin asumir propiedades estadísticas no verificables en datos *in silico* (Concordet & Haeussler, 2018; Hsu et al., 2014).

3.2.7.4. Análisis de correlación entre variables

Para explorar asociaciones monotónicas entre pares de variables cuantitativas, se empleó el coeficiente de correlación de Spearman (ρ), el cual no requiere supuestos de normalidad. Este análisis se aplicó a combinaciones relevantes de variables, incluyendo puntajes MIT versus CFD, MFE versus frecuencia de la estructura MFE, contenido GC versus eficiencia predictiva y distancia al TSS versus puntaje MIT.

Las relaciones entre variables fueron representadas mediante diagramas de dispersión, donde cada punto corresponde a una crRNA candidata, permitiendo una interpretación visual

directa de los patrones observados y su coherencia con los análisis bioinformáticos previos (Haeussler et al., 2016; Jensen et al., 2017).

3.2.7.5. Normalización de variables y análisis multicriterio

Con el fin de integrar múltiples parámetros heterogéneos en un único marco de evaluación, se implementó un análisis multicriterio ponderado. Previamente, las variables cuantitativas seleccionadas fueron normalizadas mediante escalamiento min-max, permitiendo su comparación en una escala común sin distorsionar su contribución relativa.

Posteriormente, se asignaron pesos teóricos a cada criterio en función de su relevancia en el diseño de crRNAs para CRISPRi, considerando aspectos de especificidad genómica, posicionamiento respecto al TSS y estabilidad estructural del sgRNA. El puntaje integrado se calculó como la suma ponderada de las variables normalizadas para cada crRNA, y se utilizó exclusivamente con fines de priorización racional, no como inferencia funcional definitiva (Briner et al., 2014; Li et al., 2019).

3.3. Fase 3: Diseño *in silico* del sistema CRISPRi y generación de constructos plasmídicos

3.3.1. Selección del vector base para la implementación del sistema CRISPRi

Para el diseño *in silico* del sistema CRISPRi dirigido al gen *HOXB13*, se seleccionó como vector base el plásmido lentiviral pLV hU6-sgRNA hUbC-dCas9-KRAB-T2a-GFP, originalmente reportado por Thakore et al. (2015). Este vector permite la expresión simultánea del sgRNA bajo el control del promotor humano U6 y del efector represor dCas9-KRAB bajo el promotor constitutivo hUbC, integrando además un marcador fluorescente (GFP) mediante un péptido autoclivable T2A.

La elección de este vector se fundamentó en su uso validado en aplicaciones de interferencia transcripcional, su arquitectura modular adoptada en estudios CRISPRi y la separación funcional clara entre el cassette de expresión del sgRNA y el cassette del efector

dCas9-KRAB, lo cual facilita el diseño comparativo de múltiples guías manteniendo constante el backbone plasmídico.

3.3.2. Arquitectura general del vector y orientación de los cassettes de expresión

El vector pLV hU6-sgRNA hUbc-dCas9-KRAB-T2a-GFP presenta una arquitectura bidireccional compuesta por dos cassettes de expresión independientes: El cassette hU6-sgRNA, responsable de la transcripción del sgRNA, se encuentra dispuesto en orientación reversa (reverse strand) respecto al mapa plasmídico y el cassette hUbc-dCas9-KRAB-T2a-GFP, encargado de la expresión del efector represor y del marcador fluorescente, se localiza en orientación directa (forward strand).

Esta disposición fue verificada mediante la inspección del mapa plasmídico en software SnapGene (*SnapGene software*, (www.snapgene.com)) y es consistente con la arquitectura descrita en el trabajo original de Thakore et al. (2015). La consideración explícita de la orientación de cada cassette resulta crítica para la correcta interpretación del diseño *in silico* y la documentación del ensamblaje del sistema CRISPRi.

3.3.3. Inserción *in silico* del espaciador del sgRNA en el cassette hU6

Para cada sgRNA candidata seleccionada en la Fase 2, se diseñó *in silico* un constructo derivado mediante la inserción del espaciador específico de 20 nucleótidos en el cassette hU6-sgRNA del vector base. El espaciador se posicionó aguas arriba del scaffold estándar del sgRNA, manteniendo intacta la secuencia estructural del scaffold descrita para sistemas SpCas9/dCas9.

En el mapa plasmídico del vector pLV hU6-sgRNA hUbc-dCas9-KRAB-T2a-GFP, el scaffold del sgRNA se encuentra delimitado entre las posiciones 2483–2408 pb, mientras que el espaciador variable (crRNA, 20 nt) fue insertado de forma adyacente en el intervalo 2484–2503 pb, conservando la continuidad funcional del cassette de expresión.

Dado que el cassette hU6-sgRNA se encuentra en orientación reversa, la inserción del espaciador se realizó respetando dicha orientación, garantizando la correcta transcripción del

sgARN funcional. No se introdujeron modificaciones adicionales en la longitud, composición o estructura del scaffold del sgRNA.

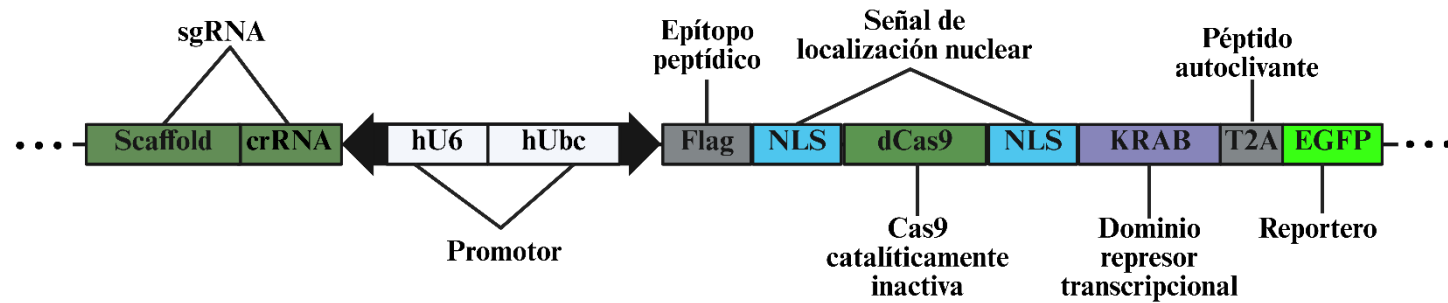
3.3.4. Conservación del cassette efector dCas9-KRAB y del marcador fluorescente

El cassette hUbc–dCas9-KRAB–T2A–GFP se mantuvo idéntico en todos los constructos diseñados, sin alteraciones en la secuencia codificante, dominios funcionales, señales de localización nuclear (NLS, *Nuclear Localization Signal*), dominio KRAB ni en el péptido T2A.

Esta estrategia metodológica permite asegurar que cualquier diferencia funcional futura entre constructos pueda atribuirse exclusivamente a la secuencia del sgRNA y no a variaciones en el backbone plasmídico, el efector represor o los elementos reguladores asociados.

3.3.5. Documentación esquemática del diseño plasmídico

La organización general del vector, la orientación de los cassettes de expresión y la localización del espaciador variable del sgRNA fueron documentadas mediante un esquema plasmídico elaborado como referencia metodológica Figura 2.



pLV hU6-sgRNA hUbC-dCas9-KRAB-T2a-GFP

Figura 2. Arquitectura del vector pLV hU6-sgRNA hUbC-dCas9-KRAB-T2a-GFP utilizada como base para el diseño *in silico* de los constructos CRISPRi. Se muestran los cassettes hU6-sgRNA (orientación inversa) y hUbC-dCas9-KRAB-T2a-GFP (orientación directa), así como la localización del espaciador variable de 20 nt insertado inmediatamente aguas arriba del scaffold del sgRNA, de acuerdo con la arquitectura descrita por Thakore et al. (2015). *Created in <https://BioRender.com>.*

3.3.6. Alcance de la Fase 3

La Fase 3 se limitó al diseño, documentación y validación conceptual *in silico* de los constructos CRISPRi, sin contemplar procedimientos experimentales de clonación, producción viral o validación funcional. Los constructos definidos constituyen la base estructural para la priorización final de sgRNAs y para una eventual implementación experimental futura del sistema CRISPRi dirigido a *HOXB13*.

CAPÍTULO 4

Resultados y discusión

4.1.1. Resultados – Fase 1: Caracterización genómica del gen *HOXB13*

4.1.1.1. Análisis en UCSC Genome Browser

El análisis realizado en el UCSC Genome Browser permitió identificar el gen *HOXB13* en el genoma de *Homo sapiens* utilizando el ensamblaje GRCh38/hg38. El gen se localiza en el cromosoma 17, banda citogenética 17q21.32, y se encuentra anotado en orientación negativa (-). A partir de las anotaciones curadas de RefSeq, se seleccionó el transcrito NM_006361.6 como referencia principal para *HOXB13*. Las coordenadas genómicas asociadas a este transcrito corresponden al intervalo chr17:48724763–48728750 (hg38). Con base en dicha anotación, se delimitó una región genómica específica correspondiente al intervalo chr17:48728450–48728800 (hg38), localizada en el entorno próximo al sitio de inicio de transcripción. La región extraída presentó un tamaño total de 351 pares de bases.

La secuencia genómica fue extraída sin aplicar enmascaramiento de repeticiones ni extensiones adicionales en los extremos 5' o 3'. El fragmento obtenido fue exportado en formato FASTA y utilizado como secuencia de entrada para los análisis posteriores de la Fase 2. Los parámetros asociados a la identificación del gen, la selección del transcrito de referencia y la delimitación de la región genómica analizada se presentan en la Tabla 1. La secuencia completa extraída se incluye en el Anexo 1.

4.1.2. Resultados-Fase 2: Diseño y preselección bioinformática de ARN guía (sgRNAs) dirigidos a *HOXB13*

4.1.2.1. Generación y preselección de crRNAs candidatos mediante CHOPCHOP

La plataforma CHOPCHOP fue utilizada para generar secuencias dirigidas a la región regulatoria del gen *HOXB13*, empleando como secuencia de entrada el fragmento genómico delimitado en la Fase 1. La configuración del análisis se realizó conforme a los parámetros

descritos previamente (Tabla 2). Como resultado, CHOPCHOP generó un conjunto inicial de 80 secuencias guía candidatas, distribuidas a lo largo de la región analizada. Estas secuencias presentaron posiciones variables respecto al sitio de inicio de transcripción (TSS) y perfiles diferenciados en los parámetros reportados por la plataforma.

Tras la aplicación del proceso de filtrado definido en la fase metodológica, el número de candidatas se redujo a nueve secuencias crRNA que cumplieron con los criterios establecidos. Estas secuencias fueron seleccionadas para continuar con los análisis posteriores de especificidad genómica y evaluación estructural.

Las secuencias retenidas fueron identificadas de acuerdo con el ranking asignado por CHOPCHOP, el cual integra métricas internas relacionadas con características de secuencia y contexto genómico. Las candidatas seleccionadas presentaron valores diferenciados en los parámetros reportados por la plataforma, lo que permitió su comparación inicial dentro del conjunto preseleccionado. Las características principales de las nueve secuencias guía seleccionadas en esta etapa se presentan en la Tabla 3, incluyendo su identificador, secuencia nucleotídica, hebra, contenido GC, autocomplementariedad y eficiencia predictiva. Este conjunto constituyó la base para los análisis desarrollados en las secciones posteriores mediante CRISPOR, Cas-OFFinder y RNAfold.

Tabla 3. Características de las crRNAs seleccionadas mediante CHOPCHOP para CRISPRi del gen *HOXB13*¹.

crRNA	Secuencia (5'–3')	Hebra	GC (%)	Autocomplementariedad	Eficiencia (%)
crRNA-1	CGGCCGTGCTGAGC GAATGC	–	70	0	47.54
crRNA-2	GCATCAGCGTAGGC GCCGCT	+	70	1	41.22
crRNA-3	CTGCGTCTCTTGCGT CAAGA	–	55	3	58.41
crRNA-4	GGGAGCGATTTAAA ACGCTT	–	45	1	42.45
crRNA-5	ATGCAGGCGACTTG CGAGCT	–	60	1	42.33
crRNA-6	TCTTGACGCAAGAG ACGCAG	+	55	1	67.12
crRNA-7	TCGCCTGCATTCGCT CAGCA	+	60	1	62.76
crRNA-10	GTCTTGACGCAAGA GACGCA	+	55	2	56.99
crRNA-12	CGTCTTGACGCAAG AGACGC	+	60	3	52.69

4.1.2.2. Determinación de coordenadas genómicas y posición relativa al TSS de las crRNAs preseleccionadas

A partir de las nueve secuencias crRNA preseleccionadas mediante CHOPCHOP (Tabla 3, Sección 4.1.2.1), se determinó su localización genómica absoluta y su posición relativa respecto al TSS del gen *HOXB13*. El cálculo se realizó siguiendo el procedimiento descrito en la Sección 3.2.2, utilizando como referencia la coordenada de inicio de la región analizada (chr17:48728450, GRCh38/hg38) y la posición del TSS del transcrito de referencia NM_006361.6 (chr17:48728750).

La conversión de la posición relativa reportada por CHOPCHOP (seq:XXX) a coordenada genómica absoluta permitió establecer, para cada crRNA, su distancia exacta al TSS. Los valores obtenidos se expresaron en pares de bases (pb), considerando valores negativos como localización aguas arriba del TSS, correspondiente a la región promotora. Los resultados del análisis posicional, incluyendo coordenadas genómicas, hebra de orientación y

¹ Se reportan las secuencias crRNA generadas por CHOPCHOP que cumplieron los criterios de preselección, incluyendo su orientación respecto a la hebra genómica, contenido GC, nivel de autocomplementariedad y eficiencia predictiva estimada por la plataforma.

distancia relativa al TSS, se presentan en la Tabla 4, permitiendo describir de forma cuantitativa la distribución de las crRNAs candidatas en el entorno regulatorio del gen *HOXB13*.

Tabla 4. Coordenadas genómicas y posición relativa al TSS de las crRNAs preseleccionadas mediante CHOPCHOP².

Rank	Secuencia crRNA	Hebra	Posición seq	Coordenada (chr17)	Distancia al TSS (pb)
1	CGGCCGTGCTGAGCGAATGC	–	265	48 728 714	–36
2	GCATCAGCGTAGGCGCCGCT	+	15	48 728 464	–286
3	CTGCGTCTCTTGCGTCAAGA	–	285	48 728 734	–16
4	GGGAGCGATTTAAAACGCTT	–	229	48 728 678	–72
5	ATGCAGGCGACTTGCGAGCT	–	249	48 728 698	–52
6	TCTTGACGCAAGAGACGCAG	+	288	48 728 737	–13
7	TCGCCTGCATTCGCTCAGCA	+	262	48 728 711	–39
10	GTCTTGACGCAAGAGACGCA	+	287	48 728 736	–14
12	CGTCTTGACGCAAGAGACGC	+	286	48 728 735	–15

4.1.2.3. Evaluación de especificidad genómica de crRNAs mediante CRISPOR

Las crRNAs preseleccionadas en la etapa de diseño inicial mediante CHOPCHOP fueron evaluadas utilizando la plataforma CRISPOR con el objetivo de obtener métricas cuantitativas de especificidad genómica y predicción de sitios de unión fuera de objetivo (off-target) en el genoma humano. El análisis se realizó de forma individual para cada crRNA candidata, empleando como referencia el genoma *Homo sapiens* (GRCh38/hg38) y el sistema SpCas9. Para cada secuencia se registraron los puntajes de especificidad MIT y CFD, así como el número total de sitios off-target predichos por la plataforma y la presencia de coincidencias con distintos niveles de desajuste.

Durante el análisis, la crRNA correspondiente al Rank 12 de CHOPCHOP fue marcada por CRISPOR con una advertencia de ausencia de correspondencia genómica. No obstante, tras la verificación manual de la secuencia y su mapeo exacto sobre el ensamblaje hg38, se confirmó su localización genómica dentro de la región analizada, por lo que se procedió a su evaluación

² Las crRNAs analizadas se localizaron exclusivamente aguas arriba del TSS del gen *HOXB13*, dentro de la región promotora, con distancias relativas comprendidas entre –13 y –286 pb. La distribución posicional evidenció la presencia de crRNAs situadas tanto en la región promotora proximal como en regiones promotoras más distales, reflejando una variabilidad espacial respecto al TSS.

bajo los mismos parámetros aplicados al resto de las crRNAs. Los resultados completos del análisis de especificidad genómica, incluyendo los puntajes MIT, CFD y el número de sitios off-target predichos para cada crRNA, se presentan en la Tabla 5.

Tabla 5. Resultados del análisis de especificidad genómica de crRNAs dirigidas al gen *HOXB13* mediante CRISPOR³.

Rank CHOPCHOP	crRNA	PAM	MIT (%)	CFD (%)	Off-targets totales
1	CGGCCGTGCTGAGCGAATGC	AGG	94	96	50
2	GCATCAGCGTAGGCCGCGCT	GGG	91	97	38
3	CTGCGTCTCTTGCGTCAAGA	CGG	95	96	43
4	GGGAGCGATTTAAAACGCTT	TGG	95	95	39
5	ATGCAGGCGACTTGCGAGCT	GGG	94	97	46
6	TCTTGACGCAAGAGACGCAG	GGG	93	94	41
7	TCGCCTGCATTCGCTCAGCA	CGG	87	94	54
10	GTCTTGACGCAAGAGACGCA	GGG	94	96	35
12	CGTCTTGACGCAAGAGACGC	AGG	94	97	26

4.1.2.4. Evaluación exhaustiva de especificidad genómica de crRNAs mediante Cas-OFFinder

Las crRNAs seleccionadas tras el análisis de especificidad en CRISPOR (Ranks 1, 3, 6, 10 y 12 de CHOPCHOP) fueron evaluadas mediante Cas-OFFinder con el fin de identificar posibles sitios fuera de objetivo en el genoma humano de referencia (*Homo sapiens*, GRCh38/hg38). Para todas las crRNAs analizadas se identificó un único sitio con coincidencia perfecta (0 mismatches), correspondiente al locus diana esperado en el cromosoma 17, asociado a la región regulatoria del gen *HOXB13*. No se detectaron sitios adicionales con 0, 1 o 2 mismatches fuera del locus objetivo para ninguna de las crRNAs evaluadas.

³ La tabla presenta los puntajes de especificidad MIT y CFD, el número total de sitios off-target predichos y la decisión de retención o descarte para cada crRNA evaluada. Las crRNAs correspondientes a los Rank 1, 3, 6, 10 y 12 fueron retenidas para las siguientes fases del análisis *in silico*.

Las diferencias entre las crRNAs se observaron en el número de sitios detectados con 3 y 4 mismatches. La crRNA Rank 1 no presentó coincidencias con 3 mismatches, mientras que las crRNAs correspondientes a los Ranks 3, 6, 10 y 12 presentaron entre uno y dos sitios en esta categoría. En relación con los sitios que presentaron 4 mismatches, todas las crRNAs mostraron múltiples coincidencias distribuidas a lo largo del genoma, con variación en el número total de eventos registrados. El número total de sitios off-target detectados para cada crRNA varió entre las candidatas analizadas, con valores comprendidos entre 26 y 50 eventos. Los perfiles completos de especificidad genómica obtenidos mediante Cas-OFFinder, desglosados por número de mismatches, se presentan en la Tabla 6.

Tabla 6. Perfil de especificidad genómica de crRNAs candidatos evaluados mediante Cas-OFFinder. ON = sitio en objetivo; OT = off-target; mm = mismatches.

Parámetro	Rank 1	Rank 3	Rank 6	Rank 10	Rank 12
crRNA	CGGCCGT GCTGAGC GAATGC	CTGCGTCT CTTGCGTC AAGA	TCTTGACGC AAGAGACG CAG	GTCTTGACG CAAGAGAC GCA	CGTCTTGAC GCAAGAGAC GC
PAM	AGG	CGG	GGG	GGG	AGG
ON-target (0 mm)	1	1	1	1	1
OT con 0–2 mm	0	0	0	0	0
OT con 3 mm	0	1	2	2	2
OT con 4 mm	50	42	37	33	24
Total, OT (≥3 mm)	50	43	39	35	26

4.1.2.5. Evaluación de la consistencia estructural del sgRNA mediante RNAfold

Con el fin de caracterizar el plegamiento secundario de las sgRNAs seleccionadas tras los análisis de especificidad genómica, se realizó un análisis estructural utilizando el servidor RNAfold. Para cada sgRNA candidata se evaluó la estructura secundaria correspondiente al sgRNA completo, considerando el espaciador de 20 nucleótidos y el scaffold estándar. El análisis generó parámetros termodinámicos cuantitativos y representaciones gráficas del paisaje conformacional del RNA, incluyendo la estructura de energía mínima libre (MFE), la estructura centroid, los mountain plots y los perfiles de probabilidad de apareamiento de bases a lo largo de la secuencia.

4.1.2.5.1. Análisis cuantitativo comparativo

Los parámetros termodinámicos obtenidos para las cinco sgRNAs evaluadas se resumen en la Tabla 7. Para cada candidata se registraron los valores de energía mínima libre (MFE), energía del ensamble termodinámico, frecuencia relativa de la estructura MFE dentro del ensamble y diversidad conformacionales del ensamble.

Los valores de MFE obtenidos se expresaron en kcal/mol y reflejan la energía asociada a la conformación estructural más estable predicha para cada sgRNA. La frecuencia de la estructura MFE se reportó como el porcentaje de ocurrencia de dicha conformación dentro del conjunto total de estructuras posibles. La diversidad conformacional se calculó como una medida global de la heterogeneidad estructural del ensamble termodinámico. Los resultados completos del análisis de plegamiento secundario y los parámetros cuantitativos asociados para cada sgRNA candidata se presentan en la Tabla 7.

Tabla 7. Parámetros termodinámicos de plegamiento secundario de crRNAs candidatas obtenidas mediante RNAfold. MFE = energía mínima libre; Frec. MFE = frecuencia de la estructura MFE dentro del ensamble; Diversidad = diversidad conformacional del ensamble⁴.

Rank CHOP CHOP	crRNA (20 nt)	MFE (kcal/mol)	Energía ensamble (kcal/mol)	Frec. MFE (%)	Diversidad
1	CGGCCGTG CTGAGCGA ATGC	-21.30	-23.59	2.44	24.66
3	CTGCGTCTC TTGCGTCAA GA	-20.90	-22.64	5.97	16.70
6	TCTTGACGC AAGAGACG CAG	-21.40	-22.98	7.68	14.99
10	GTCTTGACG CAAGAGAC GCA	-21.50	-23.06	7.90	21.54
12	CGTCTTGAC GCAAGAGA CGC	-21.20	-23.42	2.74	24.90

⁴ En conjunto, los valores de MFE se mantuvieron dentro de un rango estrecho entre candidatas, mientras que las principales diferencias se observaron en la frecuencia de la estructura MFE y en la diversidad conformacional del ensamble.

4.1.2.5.2. sgRNA Rank 1

El *mountain plot* correspondiente a la sgRNA Rank 1 mostró un pico principal localizado aproximadamente entre las posiciones 30 y 40 de la secuencia, donde las curvas asociadas a la MFE, la función de partición (*pf*) y la estructura *centroid* presentaron trayectorias coincidentes. En este mismo intervalo se registraron valores bajos de entropía estructural. En la región inicial de la secuencia (posiciones 1–20), las alturas del *mountain plot* se mantuvieron reducidas, mientras que el perfil de entropía presentó valores relativamente elevados.

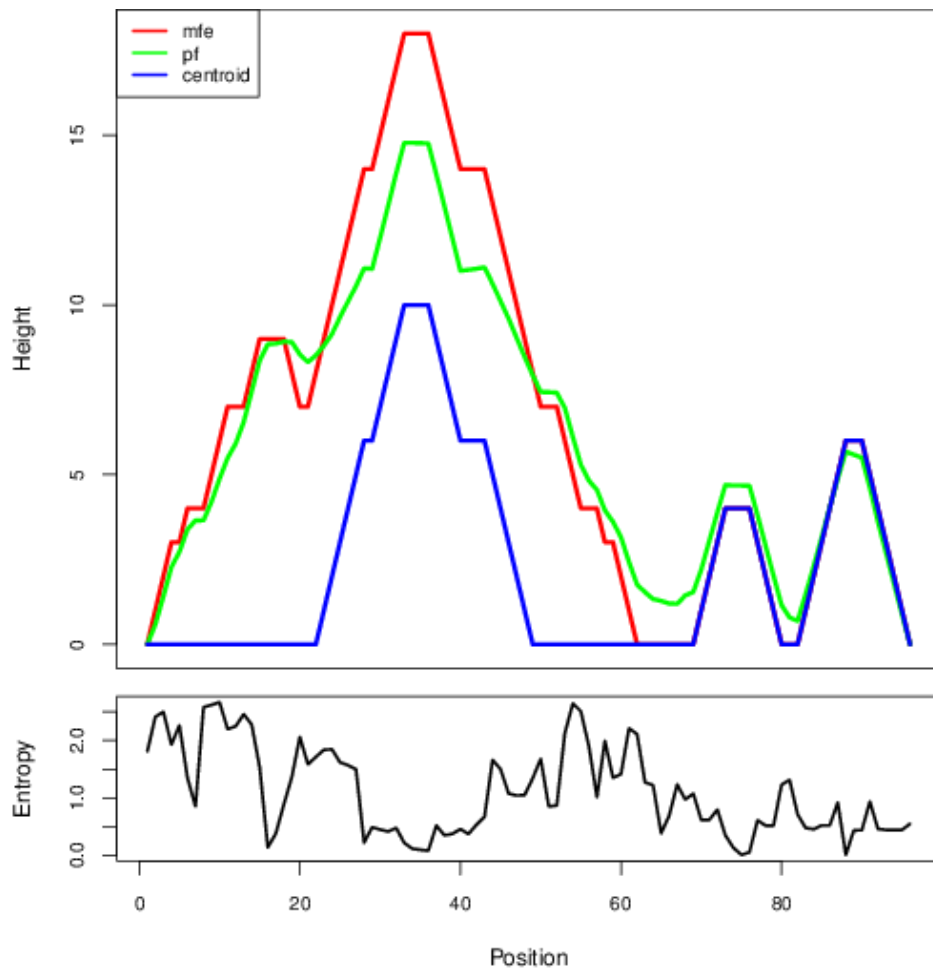


Figura 3. Mountain plot y perfil de entropía del sgRNA Rank 1. Mountain plot del sgRNA Rank 1 que muestra un pico estructural dominante entre las posiciones ~30–40, con convergencia entre las curvas MFE, ensamble termodinámico (*pf*) y *centroid*. El perfil de entropía presenta valores mínimos en esta región (<0.5) y un máximo localizado en el intervalo ~50–60, donde la altura estructural disminuye, indicando una zona de mayor flexibilidad conformacional.

El mapa de probabilidad de apareamiento de bases mostró una baja probabilidad de emparejamiento en la región correspondiente al espaciador (20 nucleótidos iniciales), seguido

de regiones con alta probabilidad de apareamiento en el scaffold del sgRNA. Se identificaron tallos definidos en la región central del scaffold, una unión de tres vías alrededor de las posiciones 49–50 y un *stem-loop* terminal conservado localizado aproximadamente entre las posiciones 80 y 96 de la secuencia.

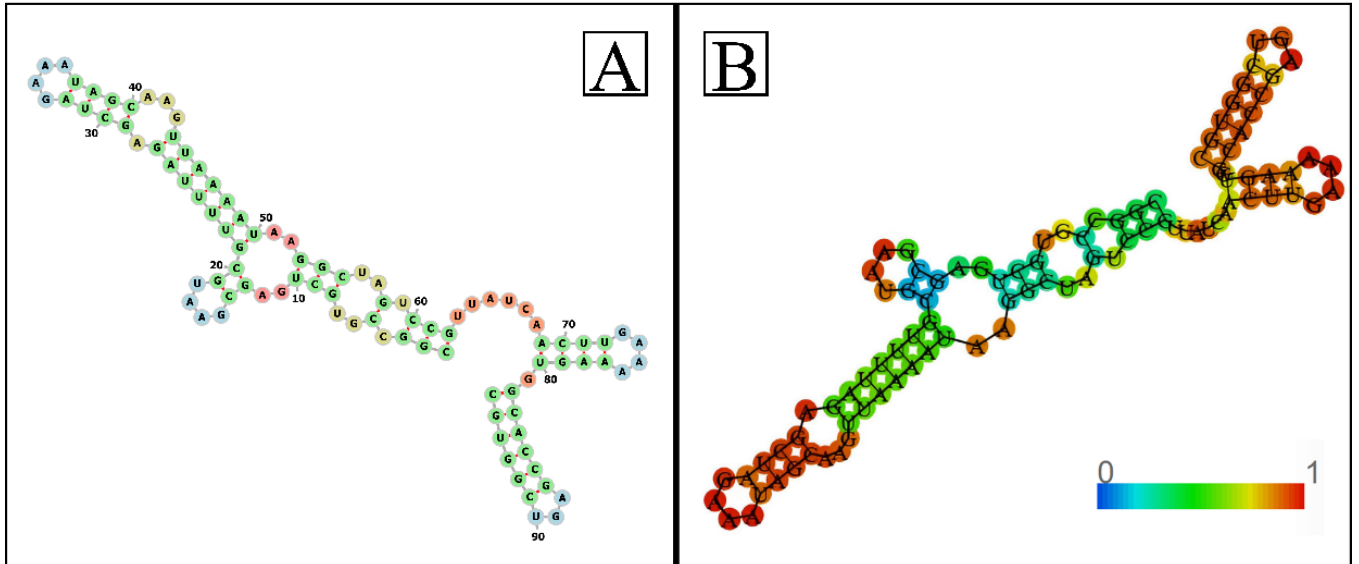


Figura 4. Análisis estructural integrado del sgRNA Rank 1 mediante RNAfold. (A) Estructura secundaria de mínima energía libre (MFE) del sgRNA Rank 1, donde se identifican los dominios estructurales canónicos del scaffold de SpCas9, incluyendo el *lower stem*, la unión de tres vías, los *hairpin stem-loops* y el *terminal stem-loop*. (B) Mapa de probabilidad de apareamiento de bases asociado a la estructura MFE, que muestra una baja probabilidad de apareamiento en la región 5' correspondiente al crRNA y altas probabilidades de apareamiento en los tallos del scaffold, especialmente en el *terminal stem-loop*. Se mide con la escala colorimétrica de probabilidad de apareamiento de bases utilizada en los análisis de RNAfold. La barra indica la probabilidad de apareamiento de bases (0–1) calculada mediante la función de partición de RNAfold. Los colores fríos (azul/turquesa) representan baja probabilidad de apareamiento, los colores intermedios (verde) probabilidad moderada y los colores cálidos (naranja/rojo) alta probabilidad de apareamiento.

4.1.2.5.3 sgRNA Rank 3

El *mountain plot* correspondiente a la sgRNA Rank 3 mostró discrepancias entre la curva asociada a la estructura de energía mínima libre (MFE) y la curva del ensamble termodinámico (*pf*) en la región central de la secuencia, aproximadamente entre las posiciones 45 y 70. En este mismo intervalo se registraron picos elevados en el perfil de entropía estructural. En contraste, en los extremos 5' y 3' de la secuencia se observó una mayor convergencia entre las curvas MFE, *pf* y *centroid*.

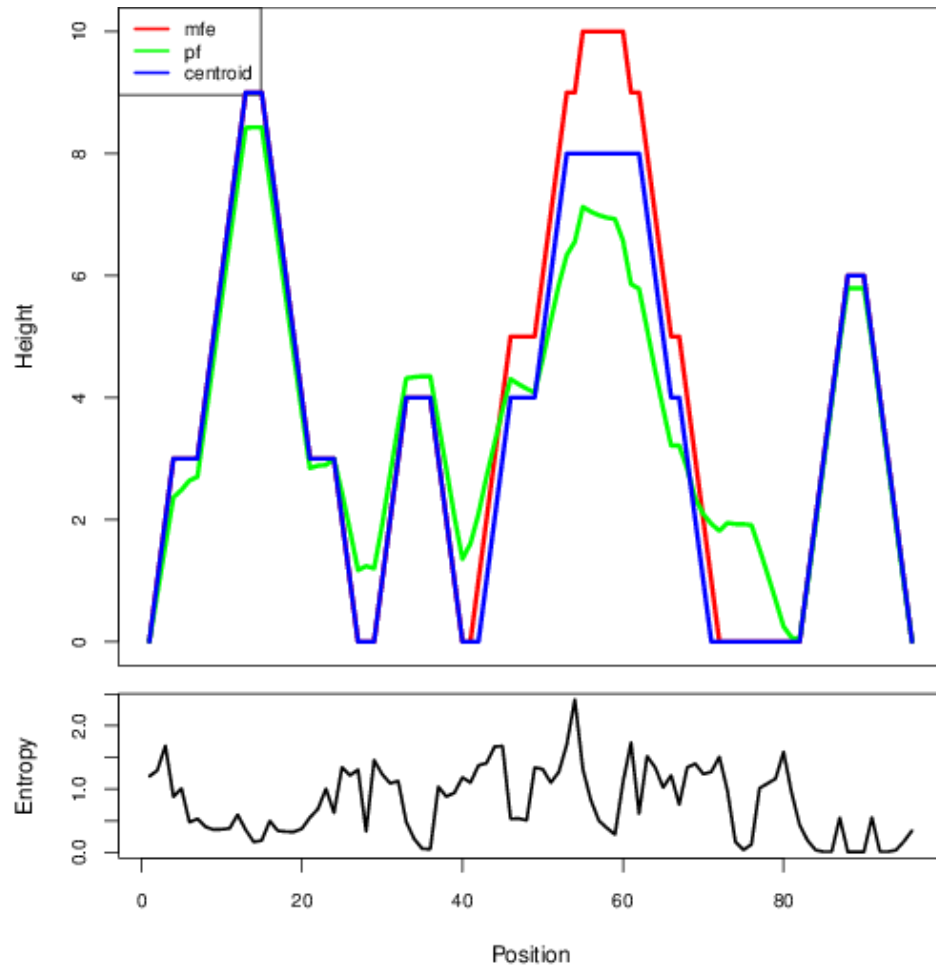


Figura 5. Mountain plot y perfil de entropía del sgRNA Rank 3. Mountain plot del sgRNA Rank 3 caracterizado por múltiples picos estructurales, incluyendo un pico principal entre las posiciones ~55–58 con divergencia visible entre las curvas MFE y pf. El perfil de entropía muestra oscilaciones frecuentes y un máximo localizado (~2.4) previo al pico estructural principal, indicando mayor dispersión conformacional en el dominio central.

El mapa de probabilidad de apareamiento de bases mostró una baja probabilidad de emparejamiento en la región correspondiente al espaciador, seguida de regiones con alta probabilidad de apareamiento en el scaffold del sgRNA. En esta región se identificaron estructuras correspondientes a la unión de tres vías y a los *stem-loops* terminales conservados.

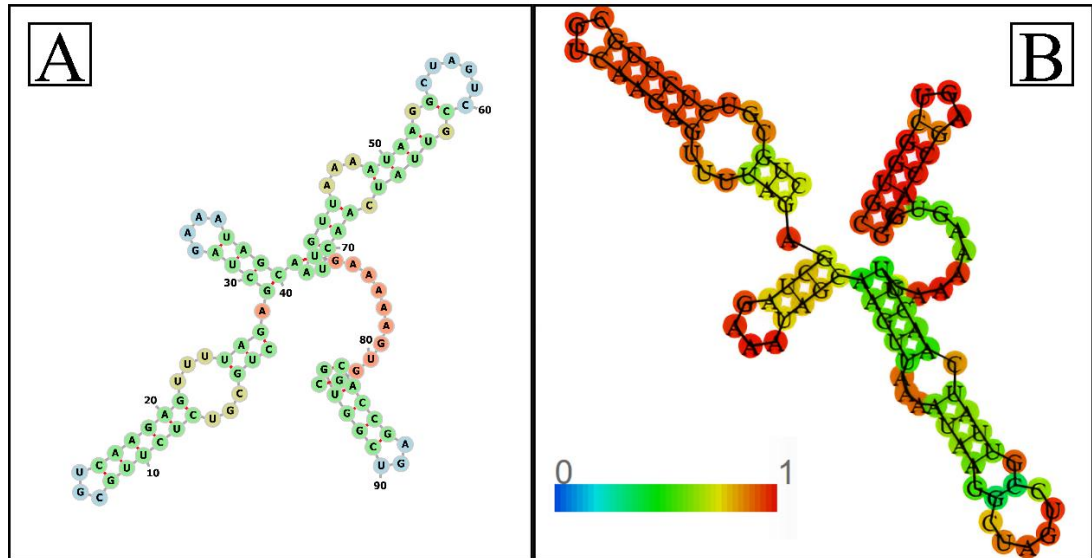


Figura 6. Análisis estructural integrado del sgRNA Rank 3 mediante RNAfold. **(A)** Estructura secundaria MFE del sgRNA Rank 3, en la que se observa un *lower stem* con un bulge asimétrico en la región basal, la conservación de la unión de tres vías y los *hairpin stem-loops* terminales. **(B)** Mapa de probabilidad de apareamiento de bases que evidencia una estabilidad intermedia en la región 5' del crRNA y probabilidades elevadas de apareamiento en los dominios del scaffold, con máxima estabilidad en el *terminal stem-loop*.

4.1.2.5.4 sgRNA Rank 6

El *mountain plot* correspondiente a la sgRNA Rank 6 mostró un pico principal pronunciado en la región central de la secuencia, aproximadamente entre las posiciones 33 y 40, donde las curvas asociadas a la MFE, al ensamble termodinámico (*pf*) y al *centroid* convergieron. En esta región se registraron valores bajos de entropía estructural. Adicionalmente, se observaron picos locales de entropía en la región inicial de la secuencia, aproximadamente entre las posiciones 5 y 15, asociados a divergencias entre la curva MFE y el ensamble termodinámico.

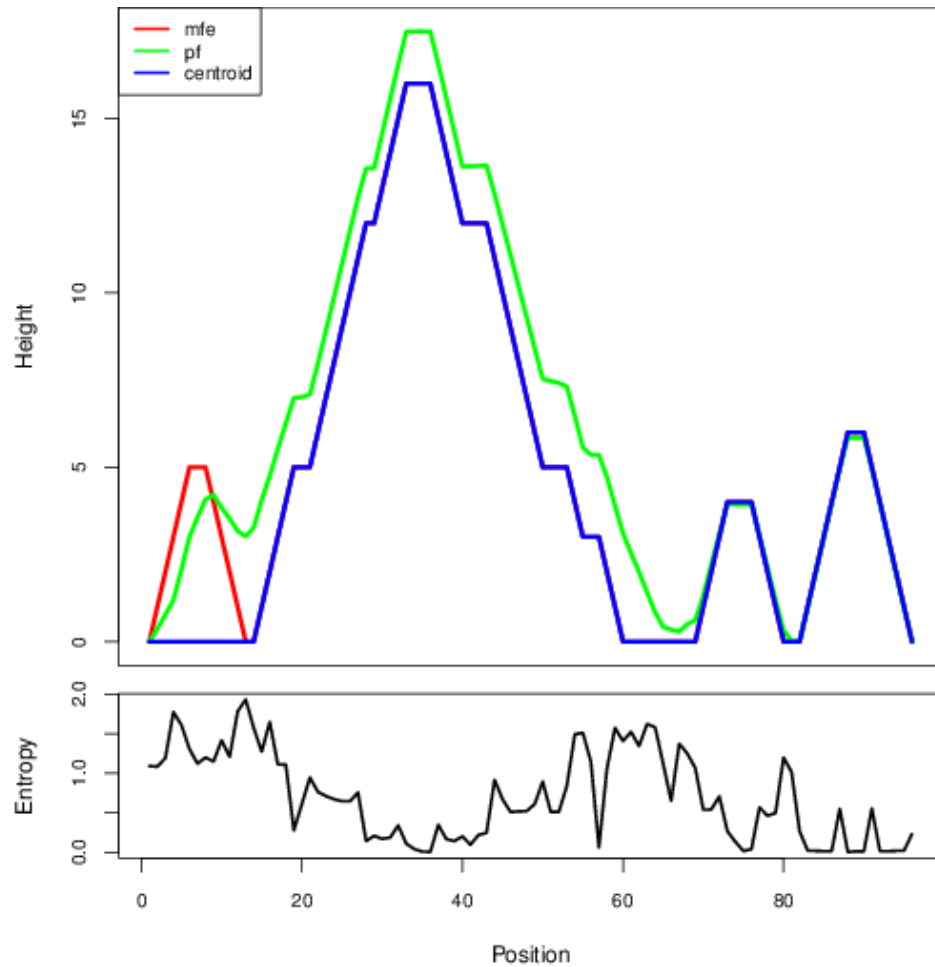


Figura 7. Mountain plot y perfil de entropía del sgRNA Rank 6. Mountain plot del sgRNA Rank 6 con un pico estructural dominante entre las posiciones ~33–40, donde las curvas MFE y centroid se solapan estrechamente. El perfil de entropía presenta valores mínimos en esta región (<0.3) y aumentos localizados en el intervalo ~55–65, coincidiendo con zonas de menor altura estructural.

El mapa de probabilidad de apareamiento de bases mostró una baja probabilidad de emparejamiento en la región correspondiente al espaciador, seguida de regiones con alta probabilidad de apareamiento en el cuerpo del sgRNA. En el extremo 3' se identificó un *stem-loop* terminal con alta probabilidad de apareamiento, consistente con las estructuras canónicas del scaffold del sgRNA.

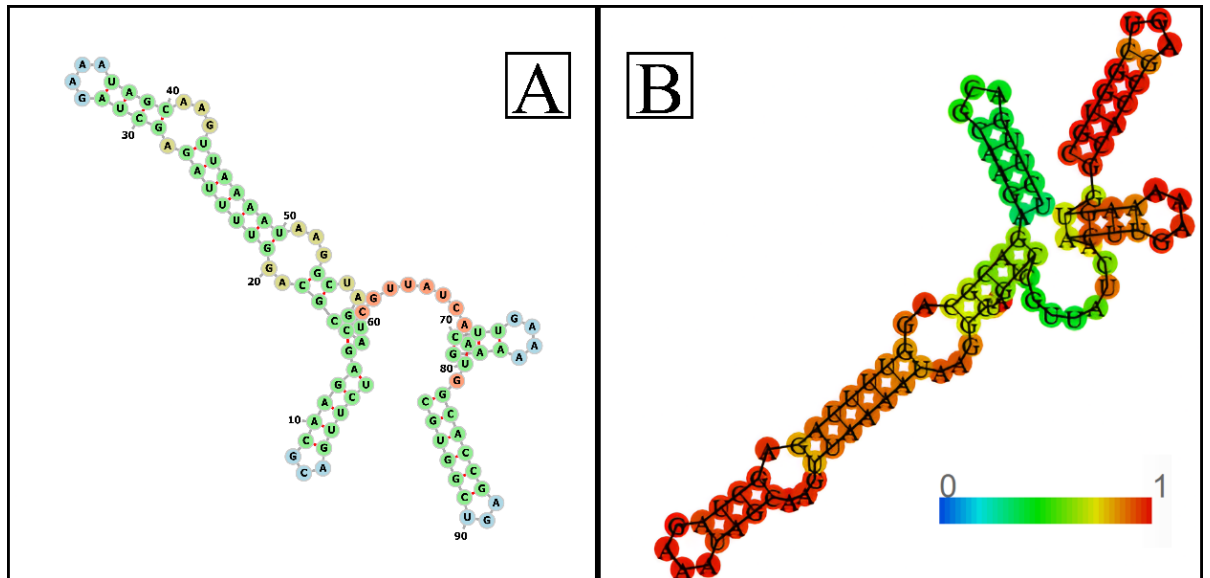


Figura 8. Análisis estructural integrado del sgRNA Rank 6 mediante RNAfold. **(A)** Estructura secundaria MFE del sgRNA Rank 6 que conserva la arquitectura canónica del scaffold de SpCas9, con un *lower stem* interrumpido por un bulge, una unión de tres vías definida y *hairpin stem-loops* terminales bien formados. **(B)** Mapa de probabilidad de apareamiento de bases que muestra alta estabilidad en los tallos centrales y terminales (colores cálidos) y menor probabilidad de apareamiento en la región 5' del crRNA.

4.1.2.5.5 sgRNA Rank 10

El *mountain plot* correspondiente a la sgRNA Rank 10 mostró un pico central definido, con transiciones graduales entre los distintos dominios estructurales. En la región inicial de la secuencia (posiciones 1–20) se registraron alturas bajas en el *mountain plot*, acompañadas de valores de entropía moderados.

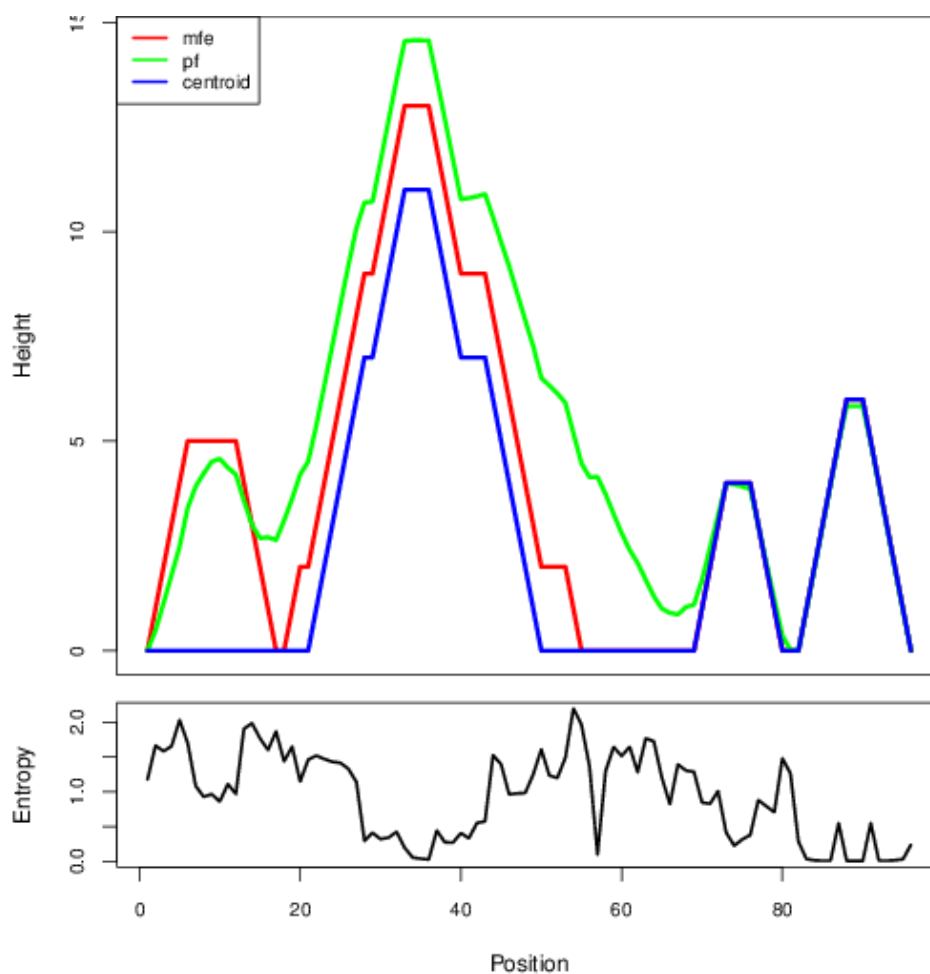


Figura 9. Mountain plot y perfil de entropía del sgRNA Rank 10. Mountain plot del sgRNA Rank 10 caracterizado por un pico estructural principal entre las posiciones ~33–37, con convergencia parcial entre MFE, pf y centroid. El perfil de entropía muestra mínimos locales asociados al pico principal y un máximo pronunciado (~2.2) en el intervalo ~50–60, indicando una región flexible entre dominios estructurales.

El mapa de probabilidad de apareamiento de bases evidenció una baja probabilidad de emparejamiento en la región correspondiente al espaciador, mientras que el scaffold presentó regiones con alta probabilidad de apareamiento. Se identificó una unión de tres vías en las posiciones A49–U50 y la presencia de *hairpins* terminales con alta probabilidad de apareamiento. En el extremo distal de la molécula, aproximadamente entre las posiciones 80 y 96, se observó una región con alta probabilidad de apareamiento de bases, consistente entre la estructura MFE, el *centroid* y el ensamble termodinámico.

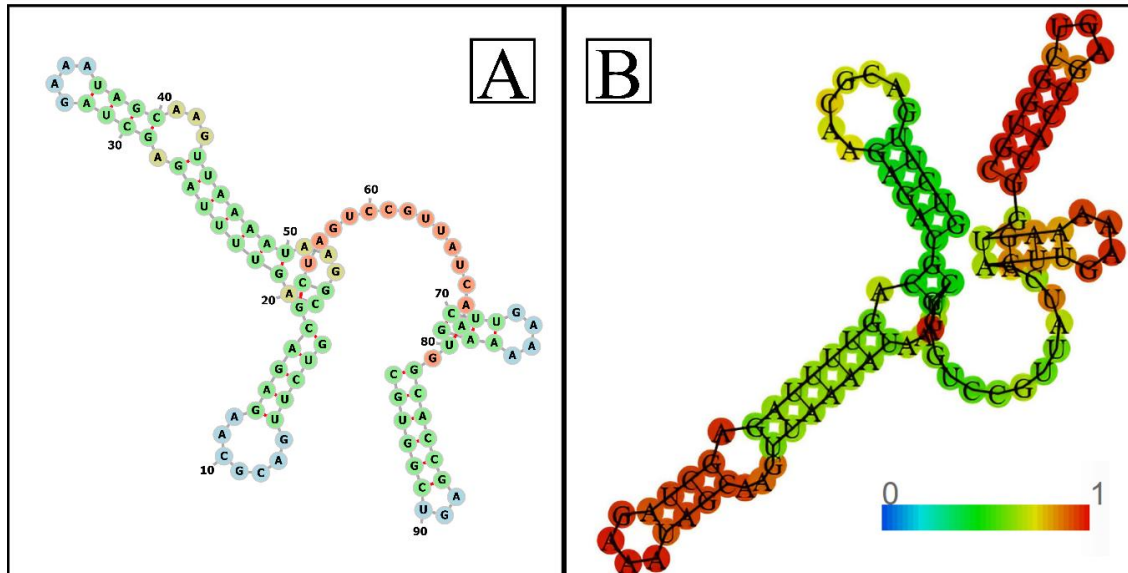


Figura 10. Análisis estructural integrado del sgRNA Rank 10 mediante RNAfold. **(A)** Estructura secundaria MFE del sgRNA Rank 10 que muestra un lower stem con bulge basal, la unión de tres vías conservada, y hairpin stem-loops bien definidos en el scaffold. **(B)** Mapa de probabilidad de apareamiento de bases donde la región 5' del crRNA presenta probabilidades bajas a moderadas de apareamiento, mientras que los dominios terminales del scaffold exhiben probabilidades altas y homogéneas.

4.1.2.5.6 sgRNA Rank 12

El *mountain plot* correspondiente a la sgRNA Rank 12 evidenció múltiples dominios estructurales, con un pico principal amplio localizado en la región central (aproximadamente posiciones 56–62) y picos secundarios definidos hacia el extremo 3' de la secuencia. En regiones intermedias (alrededor de la posición 54) se registraron valores elevados de entropía. Las curvas correspondientes a la MFE y al *centroid* convergieron en los extremos de la molécula.

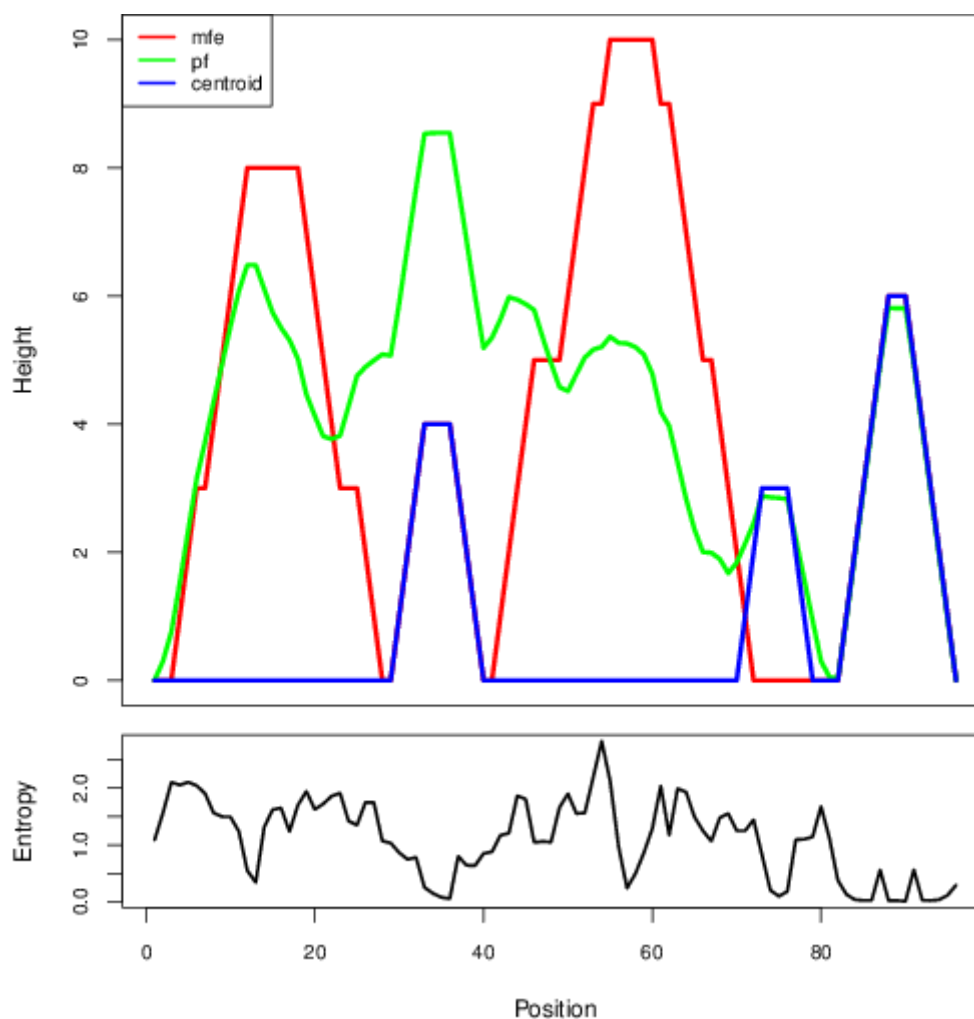


Figura 11. Mountain plot y perfil de entropía del sgRNA Rank 12. Mountain plot del sgRNA Rank 12 con una meseta estructural de altura máxima entre las posiciones ~56–62 y un pico aislado adicional en el extremo final (~88–91), donde las curvas MFE y centroid convergen. El perfil de entropía presenta un máximo localizado (~2.7) previo a la meseta principal y valores mínimos en las regiones de mayor altura estructural.

El mapa de probabilidad de apareamiento de bases mostró una baja probabilidad de emparejamiento en la región correspondiente al espaciador en el extremo 5'. A partir de esta región, se identificaron tallos con alta probabilidad de apareamiento en el scaffold del sgRNA, incluyendo un *terminal stem-loop* conservado. La región correspondiente a la unión de tres vías presentó valores intermedios de probabilidad de apareamiento.

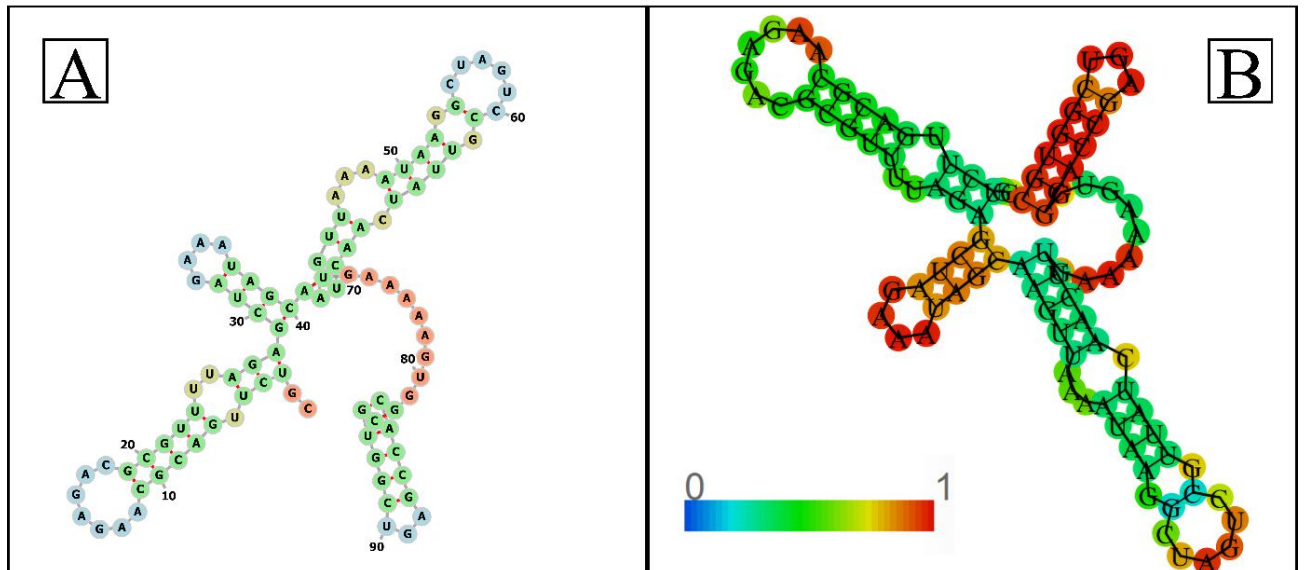


Figura 12. Análisis estructural integrado del sgRNA Rank 12 mediante RNAfold. **(A)** Estructura secundaria MFE del sgRNA Rank 12 que conserva los elementos estructurales característicos del scaffold de SpCas9, incluyendo un lower stem con bulge, la unión de tres vías y un terminal stem-loop bien definido. **(B)** Mapa de probabilidad de apareamiento de bases que muestra una baja estructuración en la región 5' del crRNA y altas probabilidades de apareamiento en los tallos terminales del scaffold.

4.1.2.6. Integración de resultados de la Fase 2

A partir de la matriz integrada generada al final de la Fase 2 (Sección 3.2.6), se realizó una síntesis comparativa de los resultados obtenidos mediante CHOPCHOP, CRISPOR, Cas-OFFinder y RNAfold para cinco sgRNAs candidatos dirigidos a la región promotora del gen *HOXB13*.

La integración consideró de forma conjunta los parámetros posicionales respecto al TSS, las métricas de especificidad genómica reportadas por CRISPOR, los perfiles exhaustivos de coincidencias off-target obtenidos mediante Cas-OFFinder y los parámetros termodinámicos derivados del análisis de plegamiento secundario del sgRNA completo mediante RNAfold. Todos los sgRNAs analizados se localizaron en la región promotora proximal del gen *HOXB13*, con distancias comprendidas entre -13 y -36 pb respecto al TSS, presentaron compatibilidad con el motivo PAM NGG y no mostraron sitios off-target con cero, uno o dos desajustes fuera del locus objetivo.

La comparación integrada de los parámetros cuantitativos evidenció diferencias entre los candidatos (Tabla 8). En particular, el número total de sitios off-target con tres o más desajustes varió entre las sgRNAs evaluadas, así como los valores asociados a energía mínima libre (MFE), frecuencia relativa de la estructura MFE y diversidad del ensamble conformacional. Asimismo, se registraron diferencias en la distribución de coincidencias off-target con tres y cuatro desajustes entre las distintas sgRNAs.

Los resultados integrados de posición genómica, especificidad y parámetros estructurales se presentan de manera comparativa en la Tabla 8, constituyendo una visión global de las características de cada sgRNA candidata derivadas del análisis *in silico* realizado en la Fase 2.

Tabla 8. Evaluación comparativa integral de cinco crRNAs candidatos para CRISPRi dirigido a *HOXB13*.

Parámetro de evaluación	Rank 1	Rank 3	Rank 6	Rank 10	Rank 12
INFORMACIÓN BÁSICA (crRNA)					
Hebra	–	–	+	+	+
Posición relativa (seq)	265	285	288	287	286
Coordenada genómica (chr17)*	48,728,714	48,728,734	48,728,737	48,728,736	48,728,735
Distancia al TSS (pb)	–36	–16	–13	–14	–15
PAM	AGG	CGG	GGG	GGG	AGG
Contenido GC (%)	70	55	55	55	60
PARÁMETROS DE DISEÑO (CHOPCHOP – crRNA)					
Rank CHOPCHOP	1	3	6	10	12
Eficiencia predicha (%)	47.54	58.41	67.12	56.99	52.69
Autocomplementariedad	0	3	1	2	3
ESPECIFICIDAD GENÓMICA (CRISPOR – crRNA)					
Puntaje MIT (%)	94	95	93	94	94
Puntaje CFD (%)	96	96	94	96	97
Off-targets totales	50	43	41	36	26
OFF-TARGETS EXHAUSTIVOS (Cas-OFFinder – crRNA)					
OT con 0 mm (fuera diana)	0	0	0	0	0
OT con 1 mm	0	0	0	0	0
OT con 2 mm	0	0	0	0	0
OT con 3 mm	0	1	2	2	2
OT con 4 mm	50	42	37	33	24
Total, OT (≥ 3 mm)	50	43	39	35	26
ESTABILIDAD ESTRUCTURAL (RNAfold – sgRNA completo)					
MFE (kcal/mol)	–21.30	–20.90	–21.40	–21.50	–21.20
Frecuencia MFE (%)	2.44	5.97	7.68	7.90	2.74
Diversidad del ensamble	24.66	16.70	14.99	21.54	24.90
Patrón del espaciador (20 nt)	(((.(((((((...((((((((((((...)))	..(((...(((..(((((((((

4.1.2.7. Resultados del análisis estadístico de la Fase 2

4.1.2.7.1. Comparación de estabilidad estructural por MFE (RNAfold)

La estabilidad estructural de las crRNAs se evaluó a partir de la energía mínima libre (MFE) calculada mediante RNAfold. La Figura 13 presenta la distribución de los valores de MFE correspondientes a cada crRNA candidata. Los valores obtenidos mostraron un rango estrecho entre las secuencias evaluadas, con variación cuantitativa entre candidatas.

El análisis estadístico global de los valores de MFE mediante la prueba no paramétrica de Kruskal–Wallis no evidenció diferencias estadísticamente significativas entre las crRNAs ($\chi^2 = 4$, $gl = 4$, $p = 0,406$).

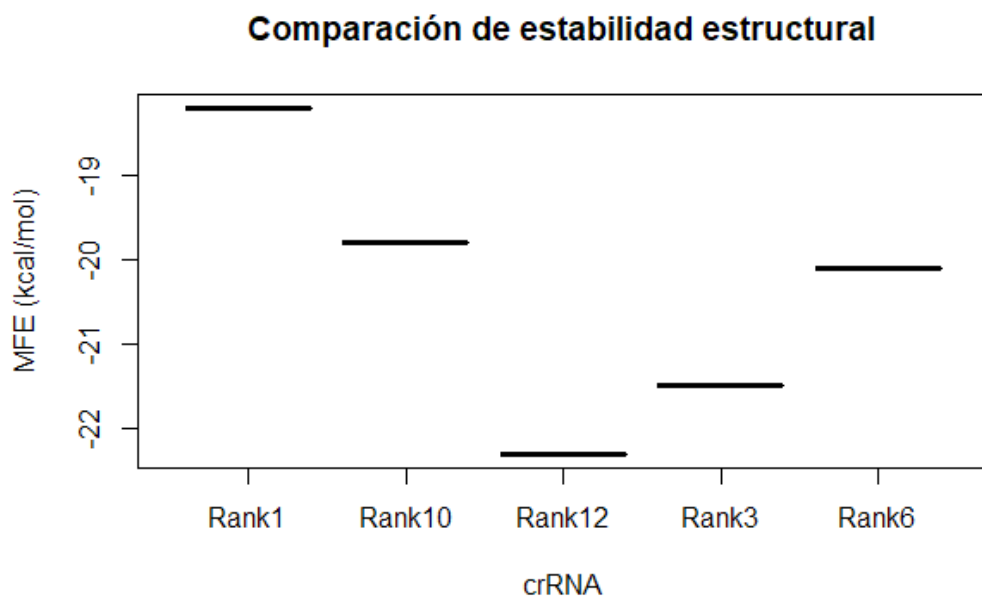


Figura 13. Comparación de estabilidad estructural por MFE. Distribución de la energía mínima libre (MFE, kcal/mol) de crRNAs candidatas obtenida con RNAfold. Valores más negativos corresponden a menor energía libre.

4.1.2.7.2. Asociación entre puntajes de especificidad MIT y CFD (CRISPOR)

La relación entre los puntajes de especificidad MIT y CFD obtenidos mediante CRISPOR para las crRNAs evaluadas se presenta en la Figura 14. Cada punto corresponde a una crRNA candidata, representando sus respectivos valores de MIT y CFD. El análisis de correlación de Spearman estimó un coeficiente $\rho = 0,50$, sin evidencia de significancia estadística ($p = 0,391$).

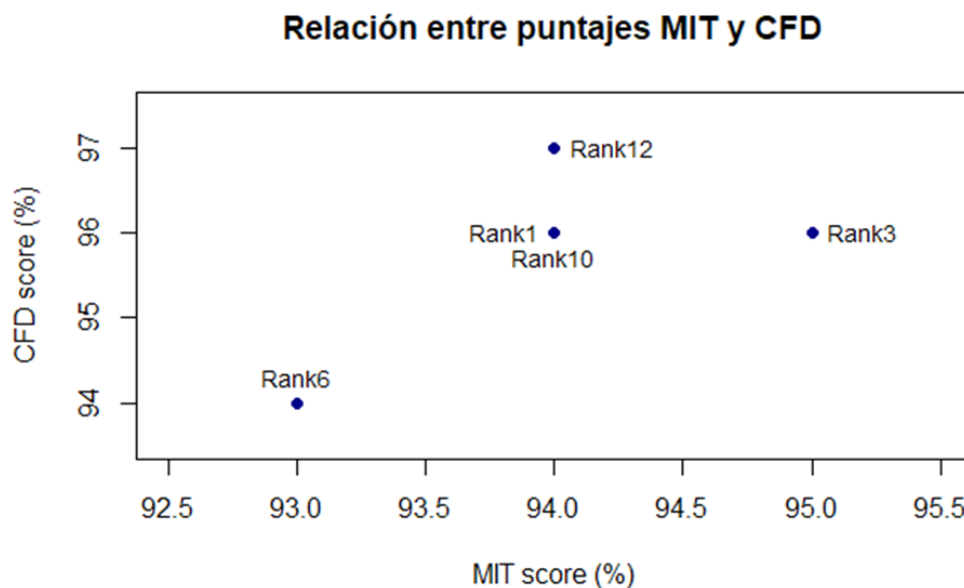


Figura 14. Relación entre puntajes MIT y CFD. Diagrama de dispersión de los puntajes de especificidad MIT y CFD para crRNAs evaluadas en CRISPOR. Cada punto corresponde a una crRNA candidata.

4.1.2.7.3. Correlaciones entre parámetros estructurales y genómicos

Se evaluaron correlaciones bivariadas entre parámetros estructurales y genómicos mediante el coeficiente de Spearman. Se obtuvo una correlación negativa entre MFE y la frecuencia de la estructura MFE ($\rho = -1$; $p = 0,0167$).

No se registraron asociaciones significativas entre contenido GC y eficiencia predictiva ($\rho = -0,447$; $p = 0,450$), ni entre distancia al TSS y puntaje MIT ($\rho = -0,671$; $p = 0,215$).

4.1.2.7.4 Ranking multicriterio de crRNAs

El análisis multicriterio ponderado generó un score integrado para cada crRNA. La Figura 15 resume el ranking final: Rank 12 presentó el score más alto (0,845), seguido por Rank 3 (0,690) y Rank 10 (0,636). Rank 6 y Rank 1 presentaron scores inferiores.

La Figura 16 muestra el mapa de calor con los valores normalizados de los parámetros incorporados en el análisis multicriterio para cada crRNA.

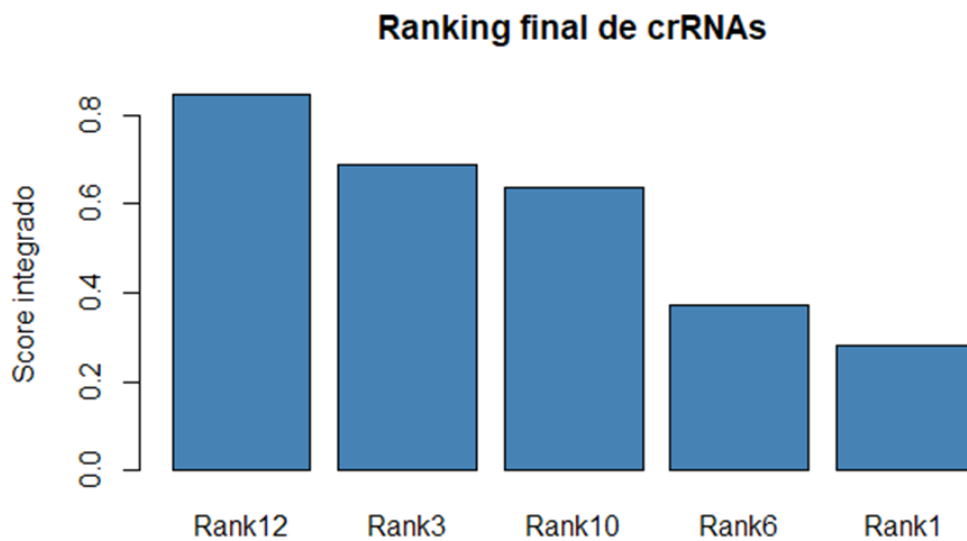


Figura 15. Ranking final de crRNAs por score integrado. Ranking de crRNAs obtenido a partir de un análisis multicriterio ponderado. El eje vertical representa el score integrado normalizado.

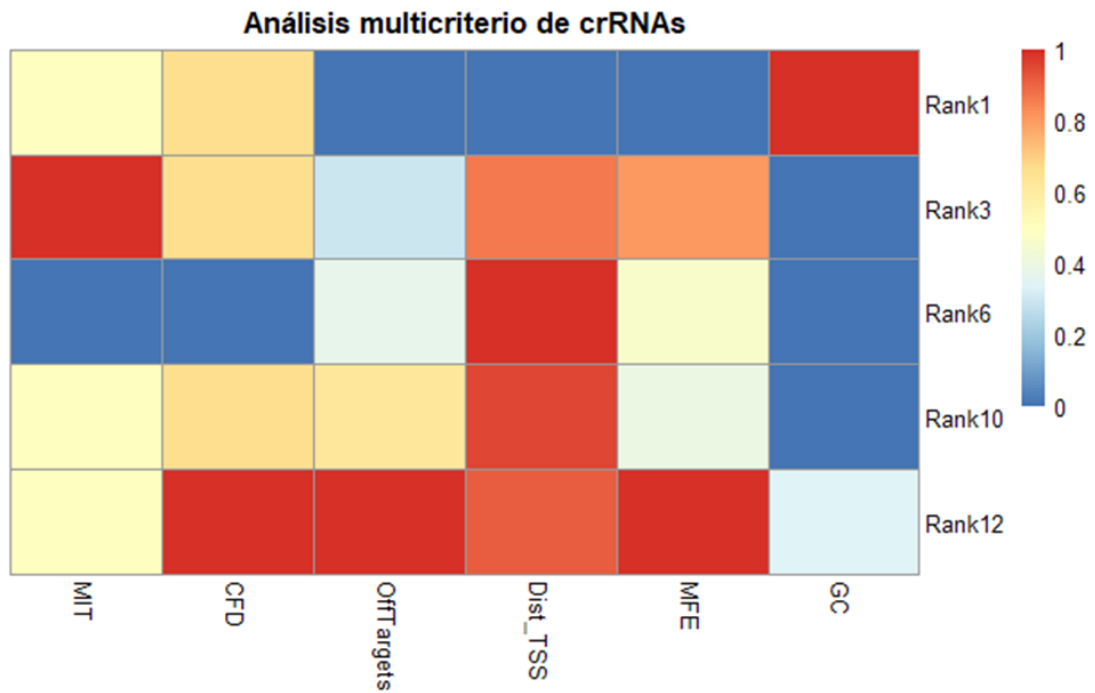


Figura 16. Mapa de calor de parámetros normalizados por crRNA. Visualización de los valores normalizados de los parámetros incorporados en el análisis multicriterio para cada crRNA candidata.

4.1.3. Resultados de la Fase 3: Generación y validación *in silico* de constructos CRISPRi

4.1.3.1. Generación de constructos *in silico* derivados del vector base

Como resultado de la Fase 3, se generaron cinco constructos *in silico* derivados del vector lentiviral pLV hU6-sgRNA hUbc-dCas9-KRAB-T2a-GFP, correspondientes a las cinco secuencias de sgRNA priorizadas tras la evaluación integral realizada en la Fase 2.

Cada constructo incorpora un espaciador específico de 20 nucleótidos en el cassette hU6-sgRNA, manteniendo inalterada la arquitectura general del vector. En todos los casos, el cassette hU6-sgRNA se localiza en orientación reversa, mientras que el cassette hUbc-dCas9-KRAB-T2A-GFP se encuentra en orientación forward, de acuerdo con el diseño original del vector.

No se detectaron alteraciones en los elementos reguladores, dominios codificantes ni secuencias funcionales del backbone plasmídico.

4.1.3.2. Documentación esquemática del constructo representativo

La organización plasmídica de los constructos generados fue documentada mediante mapas obtenidos en SnapGene. Dado que los cinco constructos difieren únicamente en la secuencia del espaciador del sgRNA, se presenta como ejemplo representativo el constructo correspondiente al Rank 12.

Alias: RANK 12

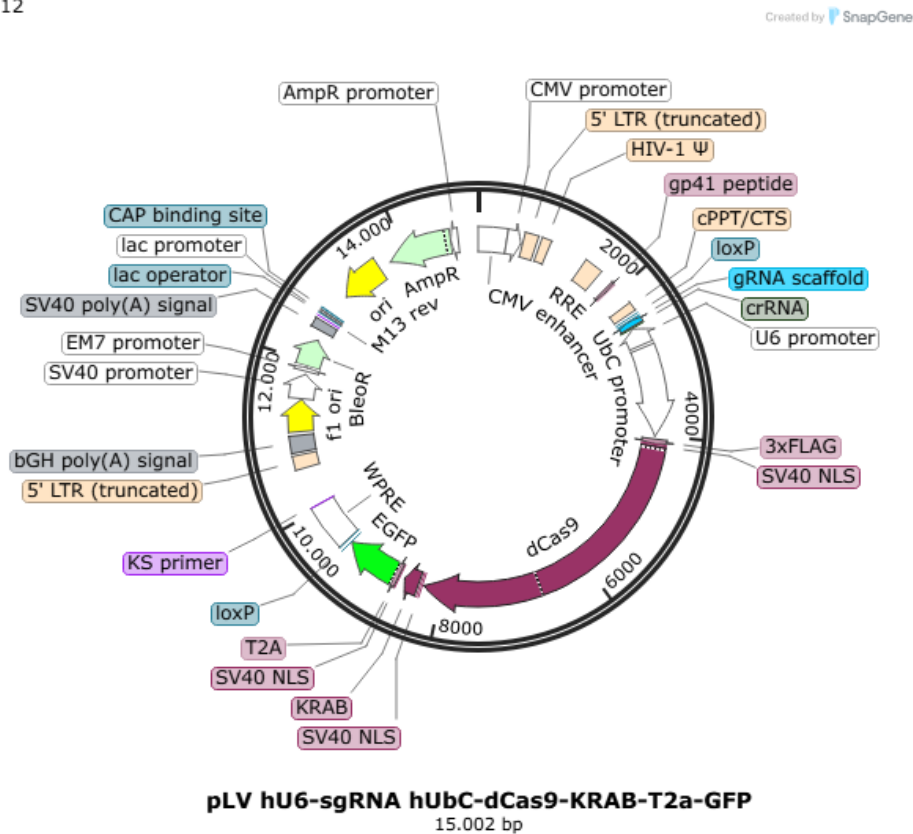


Figura 17. Mapa plasmídico circular del constructo pLV hU6-sgRNA hUbc-dCas9-KRAB-T2a-GFP con inserción del espaciador de 20 nt correspondiente al Rank 12. El esquema muestra la disposición de los cassettes de expresión y la localización del espaciador variable dentro del cassette hU6-sgRNA. Imagen obtenida de SnapGene software (www.snapgene.com).

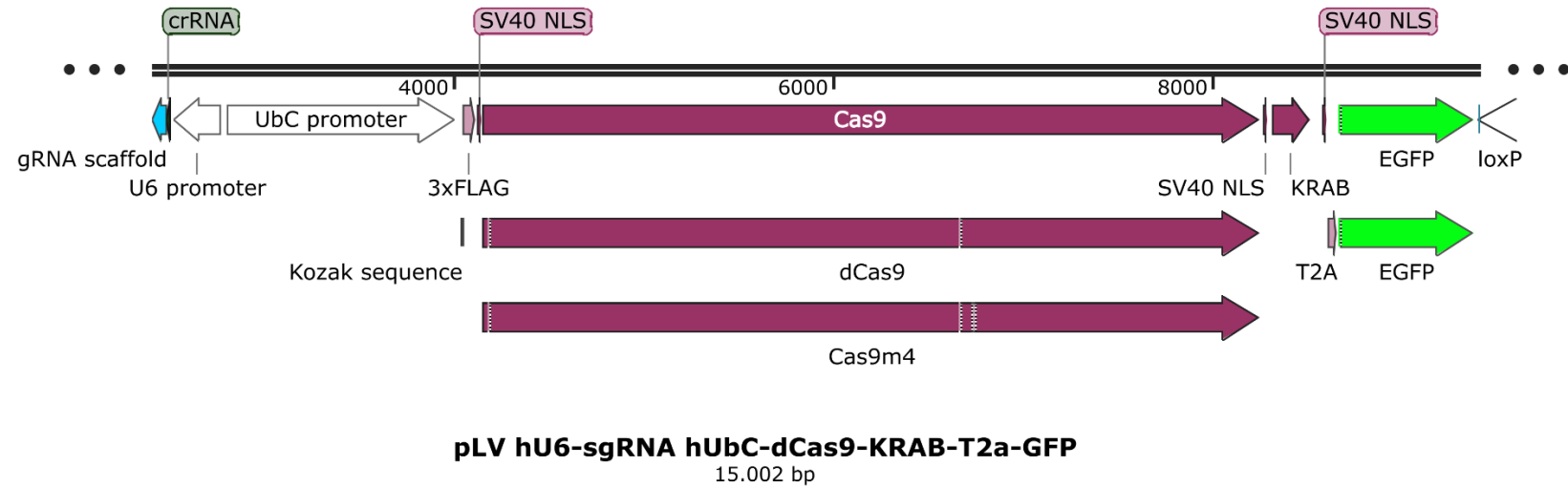


Figura 18. Mapa lineal del vector lentiviral pLV hU6-sgRNA hUbC-dCas9-KRAB-T2a-GFP (15 002 pb) generada *in silico* en SnapGene. El esquema muestra la organización del cassette hU6-sgRNA (orientación reversa), incluyendo el espaciador de 20 nt y el scaffold del sgRNA, así como el cassette hUbC-dCas9-KRAB-T2A-EGFP en orientación forward, con sus dominios funcionales anotados (3xFLAG, señales de localización nuclear SV40, dominio KRAB y gen reportero EGFP). SnapGene software (www.snapgene.com)

4.2. Discusión

El objetivo de este trabajo fue diseñar *in silico* un sistema CRISPRi dirigido al silenciamiento transcripcional del gen *HOXB13*, un regulador transcripcional central en cáncer de próstata. La relevancia de este enfoque se fundamenta en dos hechos bien establecidos: primero, *HOXB13* presenta una asociación genética robusta con cáncer de próstata hereditario y avanzado; segundo, su función como factor de transcripción lo sitúa dentro de la categoría de dianas farmacológicamente no abordables mediante inhibidores de pequeña molécula, lo que ha motivado la exploración de estrategias de interferencia a nivel génico (Barashi et al., 2024; Ewing et al., 2012; Lu et al., 2022).

Es importante señalar que los resultados obtenidos deben interpretarse considerando las limitaciones inherentes al diseño del estudio. En primer lugar, el análisis se basó en un conjunto reducido de crRNAs candidatas ($n = 5$), previamente filtradas mediante criterios bioinformáticos estrictos, lo que limita la potencia estadística para detectar diferencias globales entre guías. Adicionalmente, el carácter completamente *in silico* del estudio implica que las conclusiones derivan de modelos predictivos y métricas computacionales, sin validación experimental directa. En este contexto, la ausencia de significancia estadística ($p > 0,05$) en las pruebas aplicadas no debe interpretarse como evidencia de equivalencia funcional entre candidatas, sino como reflejo de la homogeneidad del conjunto preseleccionado y del tamaño muestral reducido. Asimismo, la presencia de empates (“ties”) en algunas métricas discretas introduce aproximaciones asintóticas en el cálculo de p-valores, lo que refuerza la necesidad de una interpretación prudente de los resultados.

En este marco, el diseño CRISPRi propuesto se orientó a maximizar la probabilidad de represión transcripcional específica mediante la integración jerárquica de criterios posicionales, genómicos y estructurales, priorizando la coherencia funcional del sistema por encima de la optimización de una métrica aislada.

La selección de sgRNAs próximas al sitio de inicio de transcripción respondió a principios funcionales descritos para CRISPRi. Gilbert et al. (2014) demostraron que la unión del complejo dCas9-KRAB en ventanas estrechas alrededor del TSS induce una represión

transcripcional más eficiente que la unión en regiones distales, al interferir con el ensamblaje de la maquinaria basal y promover la deposición de marcas represivas. De forma concordante, Thakore et al. (2015) evidenciaron que la posición relativa al TSS constituye el principal determinante de la magnitud de la represión en sistemas CRISPRi, independientemente del gen diana. En este contexto, la localización de todas las sgRNAs seleccionadas dentro de la región promotora proximal de *HOXB13* resulta coherente con el mecanismo de acción esperado del sistema, como se evidencia en las distancias relativas al TSS reportadas en la Tabla 4 y sintetizadas de manera integrada en la Tabla 8.

No obstante, la proximidad al TSS no garantiza por sí sola un silenciamiento específico. En aplicaciones CRISPRi, la unión inespecífica de dCas9-KRAB puede generar represión transcripcional no deseada sin producir señales evidentes de daño en el DNA, lo que convierte a la especificidad genómica en un criterio crítico de diseño (Gilbert et al., 2014; Listgarten et al., 2018). El análisis mediante CRISPOR y Cas-OFFinder permitió descartar la presencia de sitios off-target con cero, uno o dos desajustes fuera del locus de *HOXB13*, tal como se observa en los perfiles de especificidad genómica presentados en las Tablas 5 y 6. Las coincidencias detectadas con tres o más desajustes se alinean con lo reportado por Bae et al. (2014), quienes demostraron que la probabilidad de unión funcional disminuye de forma marcada a partir de este nivel de divergencia de secuencia. Este resultado respalda que la priorización final se construyó sobre candidatas con perfiles de especificidad comparables y previamente filtrados.

Además de la especificidad de secuencia, el plegamiento secundario del RNA guía representa un determinante funcional del sistema. Cong et al. (2013) establecieron que la integridad estructural del scaffold del sgRNA —incluyendo el *lower stem*, la unión de tres vías y los *stem-loops* terminales— es indispensable para la interacción con Cas9. Por otra parte, Konstantakos et al. (2022) y Thyme et al. (2016) señalaron que el secuestro estructural del espaciador reduce la eficiencia de hibridación con el DNA diana. En este estudio, el análisis mediante RNAfold mostró que todas las sgRNAs conservan la arquitectura canónica del scaffold y presentan baja probabilidad de apareamiento en los primeros 20 nucleótidos, como se desprende de los parámetros termodinámicos resumidos en la Tabla 7 y de los perfiles estructurales individuales presentados en las Figuras 3–12. Esta consistencia estructural también explica que, a pesar de variaciones cuantitativas en MFE, el conjunto conserve propiedades globales semejantes.

Desde una perspectiva cuantitativa, las pruebas no paramétricas de Kruskal–Wallis no evidenciaron diferencias estadísticamente significativas en los parámetros evaluados (MIT, CFD, número de off-targets, energía mínima libre y distancia al TSS; en todos los casos $\chi^2 = 4$, $gl = 4$, $p = 0,406$), lo que indica que las variaciones observadas corresponden a diferencias cuantitativas moderadas dentro de un conjunto previamente filtrado por criterios bioinformáticos estrictos. Este comportamiento es consistente con estudios de diseño de guías CRISPR donde la aplicación secuencial de filtros de eficiencia y especificidad conduce a perfiles comparables entre candidatas finales (Concordet & Haeussler, 2018; Doench et al., 2016).

No obstante, el análisis de correlación permitió identificar un patrón estructural relevante: se observó una correlación negativa significativa entre la energía mínima libre y la frecuencia relativa de la estructura MFE (Spearman $\rho = -1$; $p = 0,0167$), lo que indica que, dentro de este conjunto, las conformaciones energéticamente más favorables tienden a dominar el ensamble conformacional (Tabla 7; Figura 13). Este comportamiento es coherente con principios termodinámicos del plegamiento de RNA descritos por Lorenz et al. (2011) y ha sido asociado con una mayor estabilidad funcional del complejo CRISPR/dCas9 en estudios recientes (Riesenberg et al., 2022).

En contraste, la relación entre los puntajes MIT y CFD mostró una asociación positiva moderada, pero no significativa (Spearman $\rho = 0,50$; $p = 0,391$), como se observa en la dispersión MIT vs CFD (Figura 14), en concordancia con lo reportado por Haeussler et al. (2016). De forma complementaria, no se observaron correlaciones significativas entre contenido GC y eficiencia predictiva, ni entre distancia al TSS y puntaje MIT, lo que refuerza que estos parámetros no operan como predictores únicos dentro del rango evaluado. Cabe señalar que, en algunos contrastes, el análisis estadístico reportó advertencias por empates, por lo que los p-valores deben interpretarse como aproximaciones asintóticas en un conjunto pequeño.

Dado que los análisis univariados no permitieron discriminar de forma marcada entre candidatas, se aplicó un análisis multicriterio ponderado para integrar de manera jerárquica criterios heterogéneos, incluyendo especificidad genómica, perfil off-target, proximidad funcional al TSS, estabilidad estructural y composición de secuencia. Este enfoque ha sido

recomendado para evitar decisiones arbitrarias cuando se combinan métricas con escalas y significados distintos (Hoberecht et al., 2022; Javaid & Choi, 2021). Los resultados integrados se visualizan en el ranking final (Figura 15) y en el mapa de calor (Figura 16), donde Rank 12 emerge como la guía prioritaria al presentar el mayor *score* integrado ($\approx 0,845$), seguida por Rank 3 y Rank 10. Esta priorización debe entenderse como una hipótesis bioinformática fundamentada, no como una demostración funcional definitiva.

Desde el punto de vista vectorial, la elección del plásmido pLV hU6-sgRNA hUbc-dCas9-KRAB-T2A-GFP responde a los requisitos para su uso en células de mamífero. Dong & Kantor (2021) señalaron que la expresión coordinada del RNA guía y del efector represor, junto con señales de localización nuclear funcionales, es indispensable para la actividad CRISPRi. En este contexto, la inclusión de EGFP como marcador fluorescente permite monitorear la expresión del sistema y facilitar el seguimiento de células transducidas sin introducir presión selectiva por antibióticos, lo que resulta ventajoso en fases exploratorias (Thakore et al., 2015).

Un aspecto que relevante es la decisión de silenciar *HOXB13*, pese a su rol como regulador transcripcional. Tradicionalmente, los factores de transcripción se consideran dianas de alto riesgo debido a su participación en múltiples redes reguladoras (Norris et al., 2009). Sin embargo, estudios recientes han demostrado que *HOXB13* actúa como un nodo central oncogénico en cáncer de próstata, regulando programas transcripcionales específicos asociados a progresión tumoral, metabolismo lipídico y resistencia terapéutica (Lu et al., 2022; Xiang et al., 2025). Además, su expresión elevada y persistente en enfermedad avanzada sugiere una dependencia tumoral que justifica su modulación controlada. En este sentido, CRISPRi ofrece una ventaja conceptual frente a la edición génica permanente, al permitir una represión reversible y graduable, adecuada para estudiar reguladores transcripcionales sin eliminar de forma irreversible su función basal (Gilbert et al., 2013b; Qi et al., 2013).

Finalmente, es importante resaltar que, al tratarse de un estudio *in silico*, los resultados dependen de modelos predictivos y de un conjunto reducido de crRNAs candidatas evaluadas ($n = 5$), el cual deriva de un proceso previo de filtrado bioinformático estricto orientado a maximizar la especificidad, el posicionamiento funcional y la consistencia estructural. En este contexto, el objetivo del análisis estadístico no fue establecer inferencia poblacional ni demostrar equivalencia funcional entre candidatas, sino evaluar la coherencia interna de un

conjunto altamente preseleccionado y respaldar una priorización racional basada en múltiples criterios. En consecuencia, la ausencia de diferencias significativas ($p > 0,05$) no debe interpretarse como evidencia de equivalencia, sino como un resultado esperable tras la aplicación secuencial de filtros de diseño. Por ello, la priorización de Rank12 debe considerarse una hipótesis de trabajo, susceptible de validación experimental en etapas posteriores. En conjunto, el enfoque adoptado prioriza la integración jerárquica de parámetros genómicos, estructurales y posicionales con fines de selección informada, estableciendo hipótesis bioinformáticas explícitas que deberán ser corroboradas experimentalmente.

En conjunto, la discusión de estos resultados muestra que el sistema diseñado no constituye una simple colección de sgRNAs compatibles, sino un modelo CRISPRi construido sobre criterios funcionales, genómicos y estructurales integrados. Este enfoque se alinea con recomendaciones actuales del campo y establece una base metodológica sólida para futuras validaciones experimentales orientadas a evaluar el silenciamiento transcripcional de *HOXB13* en cáncer de próstata.

CAPÍTULO 5

Conclusiones y Recomendaciones

5.1. Conclusiones

Conclusión 1 – Objetivo específico 1

Se identificó y caracterizó una región reguladora proximal al sitio de inicio de transcripción del gen *HOXB13* en el genoma humano (GRCh38/hg38), delimitando una ventana genómica de 351 pb adecuada para el diseño de guías CRISPRi. Esta región presentó localización precisa, orientación definida y proximidad al TSS compatible con interferencia transcripcional mediada por dCas9-KRAB.

Conclusión 2 – Objetivo específico 2

El análisis *in silico* permitió seleccionar cinco sgRNAs dirigidas a la región promotora de *HOXB13* que cumplieron criterios estrictos de especificidad genómica, ausencia de off-targets críticos y consistencia estructural del RNA guía. La integración de CHOPCHOP, CRISPOR, Cas-OFFinder y RNAfold, complementada con análisis estadístico y un enfoque multicriterio ponderado, permitió priorizar las sgRNAs evaluadas, evidenciando perfiles diferenciados dentro de un conjunto homogéneo y compatible con aplicaciones CRISPRi.

Conclusión 3 – Objetivo específico 3

Se diseñaron y documentaron cinco constructos CRISPRi *in silico* derivados del vector pLV hU6-sgRNA hUbc-dCas9-KRAB-T2A-GFP, incorporando cada sgRNA seleccionada sin alterar la arquitectura plasmídica original. El sistema diseñado integra de forma coherente el cassette hU6-sgRNA y el cassette efector hUbc-dCas9-KRAB, constituyendo una plataforma genéticamente consistente y preparada para su validación funcional experimental orientada a la inhibición transcripcional de *HOXB13*.

Conclusión general

El estudio cumplió el objetivo general al diseñar *in silico* un sistema CRISPRi dirigido a inhibir la expresión del oncogén HOXB13, estableciendo un pipeline metodológico reproducible, basado en criterios genómicos, estructurales y estadísticos integrados, que permite la priorización racional de sgRNAs y sienta una base sólida para su implementación experimental futura en el contexto del cáncer de próstata.

5.2. Recomendaciones

Validar experimentalmente la sgRNA priorizada (Rank 12) mediante ensayos de represión transcripcional en líneas celulares de cáncer de próstata que expresen *HOXB13*, evaluando niveles de RNA y proteína mediante RT-qPCR y/o western blot.

Confirmar el mecanismo de represión mediado por CRISPRi mediante la evaluación de marcas epigenéticas asociadas a silenciamiento transcripcional, como H3K9me3, en regiones reguladoras del locus HOXB13.

Realizar comparaciones experimentales controladas entre las sgRNAs Rank 12 y Rank 1, considerando sus perfiles diferenciados de especificidad genómica y estabilidad estructural, con el fin de evaluar concordancia entre priorización *in silico* y desempeño funcional.

Extender el flujo de trabajo *in silico* desarrollado a otros genes reguladores implicados en cáncer de próstata, aprovechando la escalabilidad y modularidad del enfoque metodológico propuesto.

Integrar en estudios futuros datos de accesibilidad cromatínica (ATAC-seq o DNase-seq), cuando estén disponibles, para refinar la selección posicional de sgRNAs en contextos celulares específicos.

Referencias Bibliográficas

- Ahmed, M., Soares, F., Xia, J. H., Yang, Y., Li, J., Guo, H., Su, P., Tian, Y., Lee, H. J., Wang, M., Akhtar, N., Houlahan, K. E., Bosch, A., Zhou, S., Mazrooei, P., Hua, J. T., Chen, S., Petricca, J., Zeng, Y., ... He, H. H. (2021). CRISPRi screens reveal a DNA methylation-mediated 3D genome dependent causal mechanism in prostate cancer. *Nature Communications*, *12*. <https://doi.org/10.1038/s41467-021-21867-0>
- Bae, S., Park, J., & Kim, J. S. (2014). Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics*, *30*, 1473–1475. <https://doi.org/10.1093/bioinformatics/btu048>
- Barashi, N. S., Li, T., Angappulige, D. H., Zhang, B., O’Gorman, H., Nottingham, C. U., Shetty, A. S., Ippolito, J. E., Andriole, G. L., Mahajan, N. P., Kim, E. H., & Mahajan, K. (2024). Symptomatic Benign Prostatic Hyperplasia with Suppressed Epigenetic Regulator HOXB13 Shows a Lower Incidence of Prostate Cancer Development. *Cancers*, *16*. <https://doi.org/10.3390/cancers16010213>
- Bhatlekar, S., Fields, J. Z., & Boman, B. M. (2014). HOX genes and their role in the development of human cancers. In *Journal of Molecular Medicine* (Vol. 92, pp. 811–823). Springer Verlag. <https://doi.org/10.1007/s00109-014-1181-y>
- Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *74*, 229–263. <https://doi.org/10.3322/caac.21834>
- Briner, A. E., Donohoue, P. D., Gomaa, A. A., Selle, K., Slorach, E. M., Nye, C. H., Haurwitz, R. E., Beisel, C. L., May, A. P., & Barrangou, R. (2014). Guide RNA functional modules

- direct Cas9 activity and orthogonality. *Molecular Cell*, 56, 333–339. <https://doi.org/10.1016/j.molcel.2014.09.019>
- Cai, R., Lv, R., Shi, X., Yang, G., & Jin, J. (2023). CRISPR/dCas9 Tools: Epigenetic Mechanism and Application in Gene Transcriptional Regulation. *International Journal of Molecular Sciences* 2023, Vol. 24, Page 14865, 24(19), 14865. <https://doi.org/10.3390/IJMS241914865>
- Casper, J., Speir, M. L., Raney, B. J., Perez, G., Nassar, L. R., Lee, C. M., Hinrichs, A. S., Gonzalez, J. N., Fischer, C., Diekhans, M., Clawson, H., Benet-Pages, A., Barber, G. P., Vaske, C. J., van Baren, M. J., Wang, K., Rodriguez, Y. J. P., Jenkins-Kiefer, J. A., Chalamala, M., ... Haeussler, M. (2026). The UCSC Genome Browser database: 2026 update. *Nucleic Acids Research*, 54, D1331–D1335. <https://doi.org/10.1093/nar/gkaf1250>
- Chen, Y., Dufour, C. R., Han, L., Li, T., Xia, H., & Giguère, V. (2023). Hierarchical Phosphorylation of HOXB13 by mTOR Dictates Its Activity and Oncogenic Function in Prostate Cancer[Figure presented]. *Molecular Cancer Research*, 21, 1050–1063. <https://doi.org/10.1158/1541-7786.MCR-23-0086>
- Chen, Z., Wu, D., Thomas-Ahner, J. M., Lu, C., Zhao, P., Zhang, Q., Geraghty, C., Yan, P. S., Hankey, W., Sunkel, B., Cheng, X., Antonarakis, E. S., Wang, Q. E., Liu, Z., Huang, T. H. M., Jin, V. X., Clinton, S. K., Luo, J., Huang, J., & Wang, Q. (2018). Diverse AR-V7 cistromes in castration-resistant prostate cancer are governed by HoxB13. *Proceedings of the National Academy of Sciences of the United States of America*, 115, 6810–6815. <https://doi.org/10.1073/pnas.1718811115>
- Concordet, J. P., & Haeussler, M. (2018). CRISPOR: Intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Research*, 46, W242–W245. <https://doi.org/10.1093/nar/gky354>
- Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science*, 339(6121), 819–823. <https://doi.org/10.1126/science.1231143>

- Crawford, T. B., Nelson, T., Karunamuni, R., Desai, H., Hausler, R., Teerlink, C., Carter, H., Pagadala, M. S., Alba, P. R., DuVall, S. L., Danowski, M. E., Brunette, C. A., Ratner, D., Garraway, I. P., Rose, B. S., Cooney, K. A., Vassy, J. L., Hauger, R. L., Lynch, J. A., ... Maxwell, K. N. (2024). *Association of HOXB13 G84E with prostate cancer among 592,158 men*. <https://doi.org/10.1101/2024.10.15.24315450>
- Cui, Z., Huang, F., Fang, K., Yan, J., Zhang, Y., Kang, D. D., Zhou, Y., Zhao, Y., Everitt, J. I., Hankey, W., Armstrong, A. J., Huang, J., Wang, H., Jin, V. X., Dong, Y., & Wang, Q. (2025). SCORT–Cas13d Nanotherapy Precisely Targets the ‘Undruggable’ Transcription Factor HoxB13 in Metastatic Prostate Cancer In Vivo. *Advanced Science*. <https://doi.org/10.1002/advs.202417605>
- Dietlein, F., Wang, A. B., Fagre, C., Tang, A., Besselink, N. J. M., Cuppen, E., Li, C., Sunyaev, S. R., Neal, J. T., & Van Allen, E. M. (2022). Genome-wide analysis of somatic noncoding mutation patterns in cancer. *Science*, *376*. <https://doi.org/10.1126/science.abg5601>
- Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, H. W., Listgarten, J., & Root, D. E. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology*, *34*, 184–191. <https://doi.org/10.1038/nbt.3437>
- Doench, J. G., Hartenian, E., Graham, D. B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B. L., Xavier, R. J., & Root, D. E. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nature Biotechnology*, *32*, 1262–1267. <https://doi.org/10.1038/nbt.3026>
- Dong, W., & Kantor, B. (2021). Lentiviral vectors for delivery of gene-editing systems based on crispr/cas: Current state and perspectives. In *Viruses* (Vol. 13). MDPI. <https://doi.org/10.3390/v13071288>
- Dupont, W. D., Breyer, J. P., Johnson, S. H., Plummer, W. D., & Smith, J. R. (2021). Prostate cancer risk variants of the HOXB genetic locus. *Scientific Reports*, *11*. <https://doi.org/10.1038/s41598-021-89399-7>

- Erokhin, M. M., Kozelchuk, N. Y., Ziganshin, R. H., Tatarskiy, V. V., & Chetverina, D. A. (2025). HOXB13 interactome in prostate cancer cells: biochemical and functional interactions between the transcription factors HOXB13 and TBX3. *Vavilovskii Zhurnal Genetiki i Seleksii*, *29*, 744–752. <https://doi.org/10.18699/vjgb-25-82>
- Ewing, C. M., Ray, A. M., Lange, E. M., Zuhlke, K. A., Robbins, C. M., Tembe, W. D., Wiley, K. E., Isaacs, S. D., Johng, D., Wang, Y., Bizon, C., Yan, G., Gielzak, M., Partin, A. W., Shanmugam, V., Izatt, T., Sinari, S., Craig, D. W., Zheng, S. L., ... Cooney, K. A. (2012). Germline Mutations in HOXB13 and Prostate-Cancer Risk. *New England Journal of Medicine*, *366*, 141–149. <https://doi.org/10.1056/nejmoa1110000>
- Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., & Sander, J. D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology*, *31*, 822–826. <https://doi.org/10.1038/nbt.2623>
- Gilbert, L. A., Horlbeck, M. A., Adamson, B., Villalta, J. E., Chen, Y., Whitehead, E. H., Guimaraes, C., Panning, B., Ploegh, H. L., Bassik, M. C., Qi, L. S., Kampmann, M., & Weissman, J. S. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell*, *159*, 647–661. <https://doi.org/10.1016/j.cell.2014.09.029>
- Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., Stern-Ginossar, N., Brandman, O., Whitehead, E. H., Doudna, J. A., Lim, W. A., Weissman, J. S., & Qi, L. S. (2013a). XCRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, *154*, 442. <https://doi.org/10.1016/j.cell.2013.06.044>
- Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., Stern-Ginossar, N., Brandman, O., Whitehead, E. H., Doudna, J. A., Lim, W. A., Weissman, J. S., & Qi, L. S. (2013b). XCRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, *154*, 442. <https://doi.org/10.1016/j.cell.2013.06.044>
- Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R., & Hofacker, I. L. (2008). The Vienna RNA websuite. *Nucleic Acids Research*, *36*. <https://doi.org/10.1093/nar/gkn188>

- Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J. B., Schneider-Maunoury, S., Shkumatava, A., Teboul, L., Kent, J., Joly, J. S., & Concordet, J. P. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biology*, *17*. <https://doi.org/10.1186/s13059-016-1012-2>
- Hoberecht, L., Perampalam, P., Lun, A., & Fortin, J. P. (2022). A comprehensive Bioconductor ecosystem for the design of CRISPR guide RNAs across nucleases and technologies. *Nature Communications*, *13*(1), 1–20. <https://doi.org/10.1038/S41467-022-34320-7>;SUBJMETA
- Hsu, P. D., Lander, E. S., & Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. In *Cell* (Vol. 157, pp. 1262–1278). Elsevier B.V. <https://doi.org/10.1016/j.cell.2014.05.010>
- James Kent, W., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, D. (2002). The human genome browser at UCSC. *Genome Research*, *12*, 996–1006. <https://doi.org/10.1101/gr.229102>
- Javaid, N., & Choi, S. (2021). CRISPR/Cas System and Factors Affecting Its Precision and Efficiency. *Frontiers in Cell and Developmental Biology*, *9*, 761709. <https://doi.org/10.3389/FCELL.2021.761709/FULL>
- Jensen, K. T., Fløe, L., Petersen, T. S., Huang, J., Xu, F., Bolund, L., Luo, Y., & Lin, L. (2017). Chromatin accessibility and guide sequence secondary structure affect CRISPR-Cas9 gene editing efficiency. *FEBS Letters*, *591*, 1892–1901. <https://doi.org/10.1002/1873-3468.12707>
- Konstantakos, V., Nentidis, A., Krithara, A., & Paliouras, G. (2022). CRISPR–Cas9 gRNA efficiency prediction: an overview of predictive tools and the role of deep learning. *Nucleic Acids Research*, *50*(7), 3616–3637. <https://doi.org/10.1093/NAR/GKAC192>

- Labun, K., Montague, T. G., Krause, M., Torres Cleuren, Y. N., Tjeldnes, H., & Valen, E. (2019). CHOPCHOP v3: Expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Research*, *47*, W171–W174. <https://doi.org/10.1093/nar/gkz365>
- Labun, K., Rio, O., Tjeldnes, H., Swirski, M., Komisarczuk, A. Z., Haapaniemi, E., & Valen, E. (2025). CHOPOFF: symbolic alignments enable fast and sensitive CRISPR off-target detection. *BioRxiv*, 2025.01.06.603201. <https://doi.org/10.1101/2025.01.06.603201>
- Li, J., Manghwar, H., Sun, L., Wang, P., Wang, G., Sheng, H., Zhang, J., Liu, H., Qin, L., Rui, H., Li, B., Lindsey, K., Daniell, H., Jin, S., & Zhang, X. (2019). Whole genome sequencing reveals rare off-target mutations and considerable inherent genetic or/and somaclonal variations in CRISPR/Cas9-edited cotton plants. *Plant Biotechnology Journal*, *17*, 858–868. <https://doi.org/10.1111/pbi.13020>
- Listgarten, J., Weinstein, M., Kleinstiver, B. P., Sousa, A. A., Joung, J. K., Crawford, J., Gao, K., Hoang, L., Elibol, M., Doench, J. G., & Fusi, N. (2018). Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs. *Nature Biomedical Engineering*, *2*, 38–47. <https://doi.org/10.1038/s41551-017-0178-6>
- Lorenz, R., Bernhart, S. H., Höner zu Siederdisen, C., Tafer, H., Flamm, C., Stadler, P. F., & Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, *6*. <https://doi.org/10.1186/1748-7188-6-26>
- Lu, X., Fong, K. wing, Gritsina, G., Wang, F., Baca, S. C., Brea, L. T., Berchuck, J. E., Spisak, S., Ross, J., Morrissey, C., Corey, E., Chandel, N. S., Catalona, W. J., Yang, X., Freedman, M. L., Zhao, J. C., & Yu, J. (2022). HOXB13 suppresses de novo lipogenesis through HDAC3-mediated epigenetic reprogramming in prostate cancer. *Nature Genetics*, *54*, 670–683. <https://doi.org/10.1038/s41588-022-01045-8>
- Norris, J. D., Chang, C. Y., Wittmann, B. M., Kunder, R. S., Cui, H., Fan, D., Joseph, J. D., & McDonnell, D. P. (2009). The Homeodomain Protein HOXB13 Regulates the Cellular Response to Androgens. *Molecular Cell*, *36*, 405–416. <https://doi.org/10.1016/j.molcel.2009.10.020>

- Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., & Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, *152*, 1173–1183. <https://doi.org/10.1016/j.cell.2013.02.022>
- Riesenberg, S., Helmbrecht, N., Kanis, P., Maricic, T., & Pääbo, S. (2022). Improved gRNA secondary structures allow editing of target sites resistant to CRISPR-Cas9 cleavage. *Nature Communications*, *13*(1), 1–8. <https://doi.org/10.1038/S41467-022-28137-7>;TECHMETA
- Shang, Z., Zhu, S., Zhang, H., Li, L., & Niu, Y. (2013). Germline homeobox B13 (HOXB13) G84E mutation and prostate cancer risk in european descendants: A meta-analysis of 24 213 cases and 73 631 controls. In *European Urology* (Vol. 64, pp. 173–176). <https://doi.org/10.1016/j.eururo.2013.03.007>
- Thakore, P. I., D’Ippolito, A. M., Song, L., Safi, A., Shivakumar, N. K., Kabadi, A. M., Reddy, T. E., Crawford, G. E., & Gersbach, C. A. (2015). Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nature Methods*, *12*, 1143–1149. <https://doi.org/10.1038/nmeth.3630>
- Thyme, S. B., Akhmetova, L., Montague, T. G., Valen, E., & Schier, A. F. (2016). Internal guide RNA interactions interfere with Cas9-mediated cleavage. *Nature Communications*, *7*. <https://doi.org/10.1038/ncomms11750>
- Tian, Y., Dong, D., Wang, Z., Wu, L., Park, J. Y., Wei, G. H., & Wang, L. (2023). Combined CRISPRi and proteomics screening reveal a cohesin-CTCF-bound allele contributing to increased expression of RUVBL1 and prostate cancer progression. *American Journal of Human Genetics*, *110*, 1289–1303. <https://doi.org/10.1016/j.ajhg.2023.07.003>
- Xiang, R. R., Lee, S. A., Tyndall, C. F., Bhatia, A. R., Yin, J. J., Singler, C., Hauk, B. J., Kipp, M. P., & Takeda, D. Y. (2025). CRISPR screening identifies regulators of enhancer-mediated androgen receptor transcription in advanced prostate cancer. *Cell Reports*, *44*(2).

<https://doi.org/10.1016/J.CELREP.2025.115312/ASSET/ADBBCAB1-BEC8-4BC9-97DC-6751C4ABA9DB/MAIN.ASSETS/GR3.JPG>

Yang, Y., Feng, Q., Hu, K., & Cheng, F. (2021). Using CRISPRa and CRISPRi Technologies to Study the Biological Functions of ITGB5, TIMP1, and TMEM176B in Prostate Cancer Cells. *Frontiers in Molecular Biosciences*, 8. <https://doi.org/10.3389/fmolb.2021.676021>

Yi, B., Wang, S., Wang, X., Liu, Z., Zhang, C., Li, M., Gao, S., Wei, S., Bae, S., Stringer-Reasor, E., Wang, L., & Liu, R. (2022). CRISPR interference and activation of the microRNA-3662-HBP1 axis control progression of triple-negative breast cancer. *Oncogene*, 41, 268–279. <https://doi.org/10.1038/s41388-021-02089-6>

Zheng, Y., Shen, W., Zhang, J., Yang, B., Liu, Y. N., Qi, H., Yu, X., Lu, S. Y., Chen, Y., Xu, Y. Z., Li, Y., Gage, F. H., Mi, S., & Yao, J. (2018). CRISPR interference-based specific and efficient gene inactivation in the brain. *Nature Neuroscience*, 21, 447–454. <https://doi.org/10.1038/s41593-018-0077-5>

Anexos

Anexo 1. Salida Fase 1 de 351pb respecto al TSS

Gen ID (NCBI)	Región	Transcrito de referencia	Gen
10481	chr17:48728450–48728800	NM_006361.6	<i>HOXB13</i>
<p>>hg38_dna range=chr17:48728450-48728800 5' pad=0 3' pad=0 strand=+ repeat/asking=none</p> <p>ATAGTTGACAGCAGGCATCAGCGTAGGCGCCGCTGGGTGGCTGGTCAGAG GGGAGTGGGCGACCAGATTCCGCCCCCTCCCGCTCCCAGCAAGCCTTCG ATATCCTTGGCTCCATCCAAGGTGGCATAATTGCCGGGCTCCATGGAGCC GAGGGTTCGGCTCATGAGGTGCGGGGGCGGGGAATCTAGGGGGCACCCAGC TCGCTCTCCCCACCCAGGCCGGGGGAATCCAAAGCGTTTTAAATCGCTCC CAGCTCGCAAGTCGCTGCATTCGCTCAGCACGGCCGTCTTGACGCAAGA GACGCAGGGGCCCGGGCACGCGCGCTGATTGGCTGCGGCCTGGGGGAGAG</p>			

Anexo 2. Secuencia completa gen *HOXB13*

Gen ID (NCBI)	Transcrito de referencia	Longitud (bp)	Tipo de secuencia	Organismo
10481	NM_006361.6	3039	mRNA (RefSeq, MANE Select)	<i>Homo sapiens</i>
<p>>NM_006361.6 Homo sapiens homeobox B13 (<i>HOXB13</i>), mRNA</p> <p>CTCTTGCGTCAAGACGGCCGTGCTGAGCGAATGCAGGCGACTTGCGAGCTGGGAGCGATTTAAAACG CTT TGGATTCCCCCGGCCTGGGTGGGAGAGCGAGCTGGGTGCCCCCTAGATTCCCCGCCCCCGCACCT CATG AGCCGACCCTCGGCTCCATGGAGCCCGGCAATTATGCCACCTTGGATGGAGCCAAGGATATCGAAGG CTT GCTGGGAGCGGGAGGGGGGCGGAATCTGGTCGCCCACTCCCCTCTGACCAGCCACCCAGCGGCGC CTACG CTGATGCCTGCTGTCAACTATGCCCCCTTGGATCTGCCAGGCTCGGCGGAGCCGCCAAAGCAATGCC ACC CATGCCCTGGGGTGCCCCAGGGGACGTCCCCAGCTCCCGTGCCTTATGGTTACTTTGGAGGCGGGTA CTA CTCCTGCCGAGTGTCCCGGAGCTCGCTGAAACCCTGTGCCAGGCAGCCACCCTGGCCGCGTACCC CGCG GAGACTCCACGGCCGGGAAGAGTACCCAGCCGCCCACTGAGTTTGCCTTCTATCCGGGATATC</p>				

CGG
GAACCTACCAGCCTATGGCCAGTTACCTGGACGTGTCTGTGGTGCAGACTCTGGGTGCTCCTGGAGAA
CC
GCGACATGACTCCCTGTTGCCTGTGGACAGTTACCAGTCTTGGGCTCTCGCTGGTGGCTGGAACAGCC
AG
ATGTGTTGCCAGGGAGAACAGAACCCACCAGGTCCCTTTTGAAGGCAGCATTTCAGACTCCAGCGG
GC
AGCACCTCCTGACGCCTGCGCCTTTCGTGCGGGCCGCAAGAAACGCATTCCGTACAGCAAGGGGC
AGTT
GCGGGAGCTGGAGCGGGAGTATGCGGCTAACAAGTTCATACCAAGGACAAGAGGGCGCAAGATCTCG
GCA
GCCACCAGCCTCTCGGAGCGCCAGATTACCATCTGGTTTCAGAACCGCCGGGTCAAAGAGAAGAAGG
TTC
TCGCCAAGGTGAAGAACAGCGCTACCCCTAAGAGATCTCCTTGCCTGGGTGGGAGGAGCGAAAGTG
GGG
GTGTCCTGGGGAGACCAGGAACCTGCCAAGCCCAGGCTGGGGCCAAGGACTCTGCTGAGAGGCCCC
TAGA
GACAACACCCTTCCCAGGCCACTGGCTGCTGGACTGTTCTCAGGAGCGGCCTGGGTACCCAGTATG
TGC
AGGGAGACGGAACCCCATGTGACAGCCCACTCCACCAGGGTTCCCAAAGAACCTGGCCCAGTCATAA
TCA
TTCATCCTGACAGTGGCAATAATCACGATAACCAGTACTAGCTGCCATGATCGTTAGCCTCATATTTCT
ATCTAGAGCTCTGTAGAGCACTTTAGAAACCGCTTTCATGAATTGAGCTAATTATGAATAAATTTGGAAG
GCGATCCCTTTCAGGGAAAGCTTCTCTCAGACCCCTTCCATTACACCTCTCACCTTGTTAACAGCAG
G
AAGACTGAGGAGAGGGGAACGGGCAGATTCGTTGTGTGGCTGTGATGTCCGTTTAGCATTTTTCTCAGCT
GACAGCTGGGTAGGTGGACAATTGTAGAGGCTGTCTTCTCCTCCTCCTTGTCCACCCCATAGGGTGT
C
CCTACTGGTCTTGGAAAGCACCCATCCTTAATACGATGATTTTTCTGTGCGTGTGAAAATGAAGCCAGCAGGC
TGCCCTAGTCAGTCCTTCTTCCAGAGAAAAAGAGATTTGAGAAAAGTGCCTGGGTAATTCACCATTAA
TTCCTCCCCAAACTCTCTGAGTCTTCCCTTAATATTTCTGGTGGTTCTGACCAAAGCAGGTATGGTT
GTTGAGCATTGGGATCCCAGTGAAGTAGATGTTGTAGCCTTGACATACTAGCCCTTCCCAGGCACAAA
CGGAGTGGCAGAGTGGTGCCAACCCTGTTTTCCAGTCCACGTAGACAGATTACAGTGCAGGAATTCT
GG
AAGCTGGAGACAGACGGGCTCTTTCAGAGCCGGGACTCTGAGAGGGACATGAGGGCCTCTGCCTCT
GTG
TTCATTCTGATGTCCTGTACCTGGGCTCAGTGCCCGGTGGGACTCATCTCCTGGCCGCGCAGCAA
GC
CAGCGGGTTCGTGCTGGTCTTCTGACCTTAGGCTGGGGTGGGGGCCTGCCGGCGCATTCTCC
ACG
ATTGAGCGCACAGGCCTGAAGTCTGGACAACCCGAGAACCGAAGCTCCGAGCAGCGGGTCGGTGG
CGAG
TAGTGGGGTCGGTGGCGAGCAGTTGGTGGTGGGCCGCGGCCCACTACCTCGAGGACATTTCCCTC
CCG
GAGCCAGCTCTCCTAGAAACCCCGCGGGCCGCGCAGCCAAGTGTATGGCCCGCGGTCCGGT
GGGA
TCCTAGCCCTGTCTCCTCTCCTGGGAAGGAGTGAGGGTGGGACGTGACTTAGACACCTACAAATCTATT
T
ACCAAAGAGGAGCCCGGGACTGAGGGAAAAGGCCAAAGAGTGTGAGTGCATGCGGACTGGGGTTCA
GGG
GAAGAGGACGAGGAGGAGGAAGATGAGGTGATTTCTGATTTAAAAAATCGTCCAAGCCCCGTGGTC
CA

GCTTAAGGTCCTCGGTACATGCGCCGCTCAGAGCAGGTCACCTTCTGCCTCCACGTCCTCCTCAAG
 G
 AAGCCCCATGTGGGTAGCTTTCAATATCGCAGGTTCTTACTCCTCTGCCTCTATAAGCTCAAACCCACCA
 ACGATCGGGCAAGTAAACCCCTCCCTCGCCGACTTCGGAAGTGGCGAGAGTTCAGCGCAGATGGGC
 CTG
 TGGGGAGGGGGCAAGATAGATGAGGGGGAGCGGCATGGTGCGGGGTGACCCCTTGGAGAGAGGAAA
 AAGG
 CCACAAGAGGGGCTGCCACCGCCACTAACGGAGATGGCCCTGGTAGAGACCTTGGGGGTCTGGAA
 CCTC
 TGGACTCCCCATGCTCTAACTCCACACTCTGCTATCAGAACTTAACTTGAGGATTTCTCTGTTTT
 CACTCGCAATAAATTCAGAGCAAACAAA

Anexo 3. Secuencia Scaffold.

Tipo de secuencia	Sistema CRISPR	Longitud (nt)	Uso	Referencia
Scaffold (tracrRNA– crRNA fusionado)	SpCas9	80	Componente estructural del sgRNA	Cong et al. (2013)
GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCAC CGAGUCGGUGCUUUUUU				

Anexo 4. Información y secuencia del vector lentiviral CRISPRi pLV hU6-sgRNA hUbc-dCas9-KRAB-T2A-GFP.

Nombre del vector	Repositorio	Código / Plasmid ID	Aplicación	Referencia
pLV hU6- sgRNA hUbc- dCas9-KRAB- T2A-GFP	Addgene	Plasmid #71237	CRISPRi	Thakore et al. (2015)
>Addgene_NGS_Result GACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTT CCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCGCCCATGACG TCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCATTGACGTCAATGGGTGGAGTATT ACGGTAACTGCCACTTGGCAGTACATCAAGTGATCATATGCCAAGTACGCCCCCTATTGACGTCAAT GACGGTAAATGGCCCGCCTGGCATTATGCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTAC ATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGC GTTTTGACTCACGGGGATTCCAAGTCTCCACCCATTGACGTCAATGGGAGTTTGTGTTTGGCACCAAAA				

TCAACGGGACTTTCCAAAATGTCGTAACAACCTCCGCCCATGACGCAAATGGGCGGTAGGCCTGTAC
GGTGGGAGGTCTATATAAGCAGCGCCTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCC
TGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCCTGAGTGCTTCAA
GTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGA
AAATCTCTAGCAGTGGCGCCCGAACAGGGACTTGAAAAGCGAAAGGGAAACCAGAGGAGCTCTCTCGA
CGCAGGACTCGGCTTGCTGAAGCGCGCACGGCAAGAGGCGAGGGGGCGGCGACTGGTGAGTACGCC
AAAAATTTGACTAGCGGAGGCTAGAAGGAGAGAGATGGGTGCGAGAGCGTCAGTATTAAGCGGGGGA
GAATTAGATCGCGATGGGAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATAAAATTAACATATA
GTATGGGCAAGCAGGGAGCTAGAACGATTCGCAGTTAATCCTGGCCTGTTAGAAACATCAGAAGGCTGT
AGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAGATCATTATATAATA
CAGTAGCAACCCTCTATTGTGTGCATCAAAGGATAGAGATAAAAGACACCAAGGAAGCTTTAGACAAGAT
AGAGGAAGAGCAAAAACAAAAGTAAGACCACCGCACAGCAAGCGGCCGCTGATCTTCAGACCTGGAGG
AGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAGT
AGCACCCACCAAGGCAAAGAGAAGAGTGGTGCAGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTTGT
TCCTTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCGTCAATGACGCTGACGGTACAGGC
CAGACAATTATTGTCTGGTATAGTGCAGCAGCAGAACAATTGCTGAGGGCTATTGAGGCGCAACAGCAT
CTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAATCCTGGCTGTGGAAAGATACCT
AAAGGATCAACAGCTCCTGGGGATTGGGGTTGCTCTGGAAAACCTCATTTGCACCACTGCTGTGCCTTG
GAATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTTGGAATCACACGACCTGGATGGAGTGGGACAGA
GAAATTAACAATTACACAAGCTTAATACACTCCTTAATTGAAGAATCGCAAAACCAGCAAGAAAAGAATGA
ACAAGAATTATTGGAATTAGATAAATGGGCAAGTTGTGGAATTGGTTAACATAACAAATTGGCTGTGGTAT
ATAAAATTATTCATAATGATAGTAGGAGGCTTGGTAGGTTAAGAATAGTTTTTGTGTACTTTCTATAGTGAA
TAGAGTTAGGCAGGGATATTCACCATTATCGTTTCAGACCCACCTCCCAACCCCGAGGGGACCCGACA
GGCCCGAAGGAATAGAAGAAGAAGGTGGAGAGAGAGACAGAGACAGATCCATTGATTAGTGAACGGA
TCGGCACTGCGTGCGCCAATTCTGCAGACAAATGGCAGTATTCATCCACAATTTAAAAGAAAAGGGGG
GATTGGGGGGTACAGTGCAGGGGAAAGAATAGTAGACATAATAGCAACAGACATACAACTAAAGAATTA
CAAAAACAAATTACAAAATTCAAATTTTCGGGTTTATTACAGGGACAGCAGAGATCCAGTTTGGTTAATT
AATAACTTCGTATAGCATACTTATACGAAGTTATAAAAAAAGCACCGACTCGGTGCCACTTTTTCAAG
TTGATAACGGACTAGCCTATTTAACTTGCTATTTCTAGCTCTAAAACCTGAGACGTCCGTCTCCGGTGT
CGTCTTTCCACAAGATATATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTC
CATTTTAAACATAATTTTAAACTGCAAACCTACCAAGAAATATTACTTTCTACGTACAGTATTTTGTACTA
ATATCTTTGTGTTTACAGTCAAATTAATCCAATTATCTCTCTAACAGCCTTGATCGTATATGCAAATATGAA
GGAATCATGGGAAATAGGCCCTCTAATTAATTAACCCGTGTCCGGCTCCAGATCTGGCCTCCGCGCCG
GGTTTTGGCGCCTCCCGCGGGCGCCCCCTCCTCACGGCGAGCGCTGCCACGTGAGACGAAGGGC
GCAGCGAGCGTCTGATCCTTCCGCCCGGACGCTCAGGACAGCGGCCCGCTGCTCATAAGACTCGG
CCTTAGAACCCAGTATCAGCAGAAGGACATTTAGGACGGGACTTGGGTGACTCTAGGGCACTGGTTT
TCTTCCAGAGAGCGGAACAGGCGAGGAAAAGTAGTCCCTTCTCGGCGATTCTGCGGAGGGATCTCCG
TGGGGCGGTGAACGCCGATGATTATATAAGGACGCGCCGGGTGTGGCACAGCTAGTTCGGTCCGACG
CGGGATTTGGTCCGGTCTTGTGTTGTGGATCGCTGTGATCGTCACTTGGTGAGTAGCGGGCTGCTGG
GCTGGCCGGGGCTTTCGTGGCCGCGGGCCGCTCGGTGGGACGGAAGCGTGTGGAGAGACCGCCA
AGGGCTGTAGTCTGGGTCCGCGAGCAAGGTTGCCCTGAACTGGGGGTTGGGGGGAGCGCAGCAAAAT
GGCGGCTGTTCCCGAGTCTTGAATGGAAGACGCTTGTGAGGCGGGCTGTGAGGTGCTTGAACAAGGT
GGGGGGCATGGTGGGCGGCAAGAACCCAAGGCTTGGAGCCTTCGCTAATGCGGGAAAGCTCTTATTC
GGGTGAGATGGGCTGGGGCACCATCTGGGACCTGACGTGAAGTTTGTCACTGACTGGAGAACTCG
GTTTGTGCTGTTGCGGGGGCGGCAGTTATGGCGGTGCCGTTGGGCAGTGCACCCGTACCTTTGGGA
GCGCGCGCCCTCGTGTGCTGACGTCACCCGTTCTGTTGGCTTATAATGCAGGGTGGGGCCACCT
GCCGGTAGGTGTGCGGTAGGCTTTTTCTCCGTCCGAGACGCAGGGTTCCGGCCCTAGGGTAGGCTCTC
CTGAATCGACAGGCGCCGGACCTCTGGTGAGGGGAGGGATAAGTGAGGCGTCAGTTTCTTTGGTCCGT
TTTATGACCTATCTTCTAAGTAGCTGAAGCTCCGGTTTTGAACTATGCGCTCGGGGTTGGCGAGTGTG
TTTGTGAAGTTTTTAGGCACCTTTTGAATGTAATCATTGGGTCAATATGTAATTTTCAAGTGTAGACTAGTA
AATTGTCCGCTAAATTCTGGCCGTTTTTGGCTTTTTTGTAGACGAAGCTTGGGCTGCAGGTCGACTCTAG
AGCCACCATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGAT

GACAAGATGGCCCCCAAGAAGAAGAGGAAGGTGGGCCGCGGAATGGACAAGAAGTACTCCATTGGGC
TCGCCATCGGCACAAACAGCGTCGGCTGGGCCGTCATTACGGACGAGTACAAGGTGCCGAGCAAAAA
ATTCAAAGTTCTGGGCAATACCGATCGCCACAGCATAAAGAAGAACCTCATTGGCGCCCTCTGTTCTGA
CTCCGGGGAAACCGCCGAAGCCACGCGGCTCAAAAGAACAGCACGGCGCAGATATACCCGCAGAAA
GAATCGGATCTGCTACCTGCAGGAGATCTTTAGTAATGAGATGGCTAAGGTGGATGACTCTTTCTCCATA
GGCTGGAGGAGTCCTTTTTGGTGGAGGAGGATAAAAAGCACGAGCGCCACCCAATCTTTGGCAATATCG
TGGACGAGGTGGCGTACCATGAAAAGTACCCAACCATATATCATCTGAGGAAGAAGCTTGTAGACAGTA
CTGATAAGGCTGACTTGC GGTTGATCTATCTCGCGCTGGCGCATATGATCAAATTTGGGGACACTTCT
CATCGAGGGGGACCTGAACCCAGACAACAGCGATGTCGACAACTCTTTATCCAAGTGGTTCAGACTTA
CAATCAGCTTTTGAAGAGAACCCGATCAACGCATCCGGAGTTGACGCCAAAGCAATCCTGAGCGCTA
GGCTGTCCAAATCCCGGCGGCTCGAAAACCTCATCGCACAGCTCCCTGGGGAGAAGAAGAAGCGCT
GTTTGGTAATCTTATCGCCCTGTCACTCGGGCTGACCCCAACTTTAAATCTAAGTTCGACCTGGCCGAA
GATGCCAAGCTTCAACTGAGCAAAGACACCTACGATGATGATCTCGACAATCTGCTGGCCAGATCGG
CGACCAGTACGCAGACCTTTTTTGGCGGCAAAGAACCCTGTCAGACGCCATTCTGCTGAGTGATATTCTG
CGAGTGAACACGGAGATCACCAAAGCTCCGCTGAGCGCTAGTATGATCAAGCGCTATGATGAGCACCA
CCAAGACTTGACTTTGCTGAAGGCCCTTGTGAGACAGCAACTGCCTGAGAAGTACAAGGAAATTTCTTC
GATCAGTCTAAAATGGCTACGCCGGATACATTGACGGCGGAGCAAGCCAGGAGGAATTTACAAATTTA
TTAAGCCCATCTTGGAAAAATGGACGGCACCGAGGAGCTGCTGGTAAAGCTAACAGAGAAGATCTGT
TGCGCAAACAGCGCACTTTGACAATGGAAGCATCCCCACCAGATTCACCTGGGCGAACTGCACGC
TATCTCAGGCGGCAAGAGGATTTCTACCCCTTTTTGAAAGATAACAGGGAAAAGATTGAGAAAATCCTC
ACATTTGGGATAACCCTACTATGTAGGCCCCCTCGCCCGGGGAAATTCAGATTGCGGTGGATGACTCG
CAAATCAGAAGAGACCATCACTCCCTGGAACCTTCGAGGAAGTCGTGGATAAGGGGGCCTCTGCCAGT
CCTTCATCGAAAGGATGACTAATTTGATAAAAATCTGCCTAACGAAAAGTGCTTCCTAACACTCTCTG
CTGTACGAGTACTTCACAGTTATAACGAGCTCACCAAGGTCAAATACGTCACAGAAGGGATGAGAAAG
CCAGCATTCTGTCTGGAGAGCAGAAGAAAGCTATCGTGGACCTCCTCTTCAAGACGAACCGGAAAAGT
ACCGTGAACAGCTCAAAGAAGACTATTTCAAAGATTGAATGTTTCGACTCTGTTGAAATCAGCGGAGT
GGAGGATCGTTCAACGCATCCCTGGGAACGTATCACGATCTCCTGAAAATCATTAAAGACAAGGACTT
CCTGGACAATGAGGAGAACGAGGACATTCTGAGGACATTGTCCTCACCCCTACGTTGTTTGAAGATAGG
GAGATGATTGAAGAACGCTTGAAACTTACGCTCATCTCTTCGACGACAAAGTCATGAAACAGCTCAAGA
GGCGCCGATACAGGATGGGGGCGGCTGTCAAGAAAACCTGATCAATGGGATCCGAGACAAGCAGAG
TGGAAGACAATCCTGGATTTCTTAAGTCCGATGGATTTGCCAACCGGAACTTCATGCAGTTGATCCATG
ATGACTCTCTACCTTAAAGGAGGACATCCAGAAAGCACAAGTTTCTGGCCAGGGGGACAGTCTTCAG
AGCACATCGCTAATCTTGCAGGTAGCCCAGCTATCAAAAAGGGAATACTGCAGACCGTTAAGTCTGTTG
ATGAACTCGTCAAAGTAATGGGAAGGCATAAGCCCGAGAATATCGTTATCGAGATGGCCCGAGAGAACC
AACTACCCAGAAGGGACAGAAGAACAGTAGGGAAAGGATGAAGAGGATTGAAGAGGGTATAAAAGAA
CTGGGGTCCCAAATCCTTAAGGAACACCCAGTTGAAAACACCCAGCTTCAGAATGAGAAGCTCTACCTG
TACTACCTGCAGAACGGCAGGGACATGTACGTGGATCAGGAACTGGACATCAATCGGCTCTCCGACTA
CGACGTGGATGCCATCGTGCCCAAGTCTTTTCTCAAAGATGATTCTATTGATAATAAAGTGTGACAAGAT
CCGATAAAAATAGAGGGAAAGAGTGATAACGTCCCCTCAGAAGAAAGTTGTCAAGAAAATGAAAAATTATG
GCGGCAGCTGCTGAACGCCAAACTGATCACACAACCGAAGTTCGATAATCTGACTAAGGCTGAACGAG
GTGGCCTGTCTGAGTTGGATAAAGCCGGCTTCATCAAAAAGGCAGCTTGTGAGACACGCCAGATCACCA
AGCACGTGGCCCAAATTTCTGATTACAGCATGAACACCAAGTACGATGAAAATGACAACTGATTGAG
AGGTGAAAGTTATTACTCTGAAGTCTAAGCTGGTCTCAGATTTAGAAAGGACTTTTCAAGTTTATAAGGTGA
GAGAGATCAACAATTACCACCATGCGCATGATGCCTACCTGAATGCAGTGGTAGGCACTGCACCTATCA
AAAAATATCCCAAGCTTGAATCTGAATTTGTTACGGAGACTATAAAGTGTACGATGTTAGGAAAATGATCG
CAAAGTCTGAGCAGGAAATAGGCAAGGCCACCGCTAAGTACTTCTTTTACAGCAATATTATGAATTTTTTCA
AGACCGAGATTACACTGGCCAATGGAGAGATTCCGAAGCGACCACTTATCGAAACAAACGGAGAAACA
GGAGAAATCGTGTGGGACAAGGGTAGGGATTTCCGCACAGTCCGGAAGTCTGTCCATGCCGCAGG
TGAACATCGTTAAAAGACCGAAGTACAGACCGGAGGCTTCTCCAAGGAAAGTATCCTCCCGAAAAGGA
ACAGCGACAAGCTGATCGCACGCAAAAAAGATTGGGACCCCAAGAAATACGGCGGATTGATTCTCTCT
ACAGTCGCTTACAGTGTACTGGTTGTGGCCAAAGTGGAGAAAGGGAAAGTCTAAAAAACTCAAAGCGTC
AAGGAACTGCTGGGCATCACAATCATGGAGCGATCAAGCTTCGAAAAAAACCCCATCGACTTTCTCGAG

GCGAAAGGATATAAAGAGGTCAAAAAGACCTCATCATTAAAGCTTCCCAAGTACTCTCTCTTTGAGCTTGA
AAACGGCCCGAAACGAATGCTCGCTAGTGCGGGCGAGCTGCAGAAAGGTAACGAGCTGGCACTGCC
CTCTAAATACGTTAATTTCTGTATCTGGCCAGCCACTATGAAAAGCTCAAAGGGTCTCCCGAAGATAATG
AGCAGAAGCAGCTGTTTCGTGGAACAACACAACACTACCTTGATGAGATCATCGAGCAAATAAGCGAAT
TCTCCAAAAGAGTGATCCTCGCCGACGCTAACCTCGATAAGGTGCTTTCTGCTTACAATAAGCACAGGG
ATAAGCCCATCAGGGAGCAGGCAGAAAACATTATCCACTTGTTACTCTGACCAACTTGGGCGCGCCTG
CAGCCTTCAAGTACTTCGACACCACCATAGACAGAAAGCGGTACACCTCTACAAAGGAGGTCCTGGAC
GCCACACTGATTCATCAGTCAATTACGGGGCTCTATGAAAACAAGAATCGACCTCTCTCAGCTCGGTGGA
GACAGCAGGGCTGACCCCAAGAAGAAGAGGAAGGTGGCTAGCGATGCTAAGTCACTGACTGCCTGGT
CCCGGACACTGGTGACCTCAAGGATGTGTTTGTGGACTTACCAGGGAGGAGTGGAAGCTGCTGGAC
ACTGCTCAGCAGATCCTGTACAGAAATGTGATGCTGGAGAACTATAAGAACCCTGGTTTCTTGGGTATCA
GCTTACTAAGCCAGATGTGATCCTCCGGTTGGAGAAGGGAGAAGAGCCCTGGCTGGTGGAGAGAGAAA
TTCACCAAGAGACCCATCCTGATTCAGAGACTGCATTTGAAATCAAATCATCAGTTCGGAAAAAGAAACG
CAAAGTTGCTAGCGAGGGCAGAGGAAGTCTTCTAACATGCGGTGACGTGGAGGAGAATCCCGGCCCT
GGTACCGTGAGCAAGGGCGAGGAGCTGTTACCCGGGGTGGTGCCCATCCTGGTTCGAGCTGGACGGC
GACGTA AACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG
ACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGA
CCTACGGCGTGACGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGC
CATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGC
GCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGG
AGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCCACAACGTCTATATCATGGCC
GACAAGCAGAAGAACGGCATCAAGGTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGC
AGCTCGCCGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAAC
CACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTG
TGGAGTTCGTGACCGCCGCGGGATCACTCTCGGCATGGACGAGCTGTACAAGACCGGTTGATAATAG
ATAACTTCGTATAGCATAATTATACGAAGTTATGAATTCGATATCAAGCTTATCGATAATCAACCTCTGGAT
TACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTT
TAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTTCAATTTCTCCTCCTTGTATAATCCTGGTTGCT
GTCTCTTTATGAGGAGTTGTGGCCCGTTGTGAGGCAACGTGGCGTGGTGTGCACTGTGTTTGTGACGC
AACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTGCTTTCCCCCTCC
CTATTGCCACGGCGAACTCATCGCCGCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGG
GCACTGACAATCCGTGGTGTGTCGGGGAAATCATCGTCTTTCCCTGGCTGCTCGCCTGTGTTGCCA
CCTGGATTCTGCGCGGGACGTCTTCTGCTACGTCCCTTCCGGCCCTCAATCCAGCGGACCTTCTTCC
CGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCCGCTTCCGCCCTCAGACGAGTCGGATCT
CCCTTTGGGCCGCTCCCCGCATCGATACCGTCGACCTCGAGACCTAGAAAAACATGGAGCAATCAC
AAGTAGCAATACAGCAGCTACCAATGCTGATTGTGCTGGCTAGAAGCACAAGAGGAGGAGGAGGTGG
GTTTTCCAGTCACACCTCAGGTACCTTAAAGACCAATGACTTACAAGGCAGCTGTAGATCTTAGCCACTTT
TAAAAGAAAAGGGGGGACTGGAAGGGCTAATCACTCCAACGAAGACAAGATATCCTTGATCTGTGG
ATCTACCACACACAAGGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGATCAGATATCC
ACTGACCTTTGGATGGTGTACAAGCTAGTACCAGTTGAGCAAGAGAAGGTAGAAGAAGCCAATGAAGG
AGAGAACACCCGCTTGTTACACCCTGTGAGCCTGCATGGGATGGATGACCCGGAGAGAGAAGTATTAG
AGTGGAGGTTTGACAGCCGCCTAGCATTTCATCACATGGCCGAGAGCTGCATCCGGACTGTACTGGG
TCTCTGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAAGCCT
CAATAAGCTTGCCTTGAGTGTCTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATC
CCTCAGACCCTTTAGTCAGTGTGAAAATCTCTAGCAGGGCCCGTTTAAACCCGCTGATCAGCCTCGA
CTGTGCCCTTCTAGTTGCCAGCCATCTGTTGTTGCCCTCCCCCGTGCCTTCTTACCCTGGAAGGTG
CCACTCCCACTGTCTTTCCTAATAAAAATGAGGAAATGCATCGCATTGTCTGAGTAGGTGTCAATTCTATT
TGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGG
ATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGC
GCCCTGTAGCGGCGCATTAAAGCGCGGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGC
CAGCGCCCTAGCGCCGCTCCTTTCGCTTCTTCCCTTCTTCTCGCCACGTTCCCGGCTTCCCC
GTCAAGCTCTAAATCGGGGGCTCCCTTAGGGTCCGATTAAGTGCTTACGGCACCTCGACCCAAAA

AACTTGATTAGGGTGATGGTTCACGTAAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTT
GGAGTCCACGTTCTTAATAGTGGACTCTTGTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATT
CTTTGATTTATAAGGGATTTTCCGATTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTAAC
GCGAATTAATTCTGTGGAATGTGTGTCAGTTAGGGTGTGAAAAGTCCCCAGGCTCCCCAGCAGGCAGAA
GTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGAAAAGTCCCCAGGCTCCCCAGCAGGC
AGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCC
GCCCTAACTCCGCCAGTTCGCCATTCTCCGCCCATGGCTGACTAATTTTTTTATTTATGCAGAG
GCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTT
TGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCTGATCAGCACGTGTTGACAATTAATCATCG
GCATAGTATATCGGCATAGTATAATACGACAAGGTGAGGAACTAAACCATGGCCAAGTTGACCAGTGCC
GTTCCGGTGCTCACCGC CGCGACGTCGCCGGAGCGGTGAGTTCTGGACCGACCGGCTCGGGTTC
TCCCGGACTTCGTGGAGACGACTTCGCCGGTGTGGTCCGGGACGACGTGACCCTGTTTCATCAGCG
CGTCCAGGACCAGGTGGTGCCGGACAACACCCTGGCCTGGGTGTGGGTGCGCGCCCTGGACGAG
CTGTACGCCGAGTGGTCGGAGGTCGTGTCCACGAACTTCGGGACGCCTCCGGGCCGGCCATGACC
GAGATCGGCCGAGCAGCCGTGGGGCGGGAGTTCGCCCTGCGCGACCCGGCCGGCAACTGCGTGCA
CTTCGTGGCCGAGGAGCAGGACTGACACGTGCTACGAGATTCGATTCCACCGCCGCCTTCTATGAAA
GGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTG
GAGTCTTCGCCACCCCAACTTGTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAAT
TTCACAAATAAAGCATTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAACTCATCAATGTATCTTATCATGT
CTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCTGTGTGAAATTGTT
ATCCGCTCACAAATCCACACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAATGAG
TGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTCCAGTCGGGAAACCTGTGCTGCCAGC
TGCATTAATGAATCGCCAACGCGCGGGGAGAGGCGTTTGCATTTGGGCGCTCTCCGCTTCTCCTG
CTCACTGACTCGCTGCGCTCGGTCTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAA
TACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAAGGCCAGCAAAAAGGCC
AGGAACCGTAAAAAGGCCGCGTGTGCGGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAA
AAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTG
GAAGCTCCCTCGTGCCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTCTCCCT
TCGGGAAGCGTGGCGCTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTGCTTCGCTCC
AAGCTGGGCTGTGTGCACGAACCCCCGTTCCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCT
TGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAG
CGAGGTATGTAGGCGGTGCTACAGAGTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAG
TATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAA
ACAAACCACCGCTGGTAGCGGTGGTTTTTTGTTTGAAGCAGCAGATTACGCGCAGAAAAAAGGATCT
CAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGAACGAAAACTCACGTTAAGGGATTTT
GGTCATGAGATTATCAAAAAGGATCTTACCTAGATCCTTTAAATTAATAAATGAAGTTTTAAATCAATCTAAA
GTATATATGAGTAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGT
CTATTTGTTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCAT
CTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAAC
CAGCCAGCCGGAAGGGCCGAGCGCAGAAAGTGGTCTGCAACTTTATCCGCCTCCATCCAGTCTATTAA
TTGTTGCCGGGAAGCTAGAGTAAGTAGTTCCGCAAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACA
GGCATCGTGGTGTACGCTCGTCTGTTGGTATGGCTTCATTACGCTCCGGTTCCCAACGATCAAGGCCGA
GTTACATGATCCCCATGTTGTGCAAAAAGCGGTTAGCTCCTTCGGTCTCCGATCGTTGTCAGAAGTA
AGTTGGCCGCAAGTGTATCACTCATGGTTATGGCAGCACTGCATAATTCTTACTGTCATGCCATCCGTA
AGATGCTTTTCTGTGACTGGTGAAGTCAACCAAGTCACTTCTGAGAATAGTGTATGCGGCGACCGAGTT
GCTCTTGCCCGGCGTCAATACGGGATAATACCGGCCACATAGCAGAACTTTAAAAGTGCTCATCATTG
GAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCA
CTCGTGCACCCAACTGATCTTACGATCTTTACTTTACCAGCGTTTCTGGGTGAGCAAAAACAGGAAG
GCAAAATGCCGCAAAAAGGGAATAAGGGCCGACACGGAAATGTTGAATACTCATACTCTTCTTTTCAAT
ATTATTGAAGCATTATCAGGGTATTGTCTCATGAGCGGATACATATTTGAATGATTTAGAAAAATAACAA
ATAGGGGTTCCGCGCACATTTCCCGAAAAGTGCCACCTGACGTCGACGGATCGGGAGATCTCCCGAT
CCCCTATGGTGCCTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTT

GTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAATTTAAGCTACAACAAGGCAAGGCTGACCGACA
ATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGC
GTT

Anexo 5. Secuencias de sgRNAs CRISPRi dirigidas a HOXB13 (crRNA + scaffold estándar)

Rank (CHOPCHOP)	crRNA (20 nt)	Scaffold CRISPR (Cong et al., 2013)	Secuencia sgRNA completa (5'→3')
Rank 1	CGGCCG TGCTGA GCGAAT GC	GUUUUAGAGCUAGAAAU AGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACU UGAAAAAGUGGCACCGA GUCGGUGCUUUUUU	CGGCCGTGCTGAGCGAA TGCGUUUUAGAGCUAG AAAUAGCAAGUUAAAA UAAGGCUAGUCCGUUA UCAACUUGAAAAAGUG GCACCGAGUCGGUGCU UUUUU
Rank 3	CTGCGT CTCTTG CGTCAA GA	GUUUUAGAGCUAGAAAU AGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACU UGAAAAAGUGGCACCGA GUCGGUGCUUUUUU	CTGCGTCTCTTGCGTCA AGAGUUUUAGAGCUAG AAAUAGCAAGUUAAAA UAAGGCUAGUCCGUUA UCAACUUGAAAAAGUG GCACCGAGUCGGUGCU UUUUU
Rank 6	TCTTGA CGCAA GAGAC GCAG	GUUUUAGAGCUAGAAAU AGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACU UGAAAAAGUGGCACCGA GUCGGUGCUUUUUU	TCTTGACGCAAGAGACG CAGGUUUUAGAGCUAG AAAUAGCAAGUUAAAA UAAGGCUAGUCCGUUA UCAACUUGAAAAAGUG GCACCGAGUCGGUGCU UUUUU
Rank 10	GTCTTG ACGCA AGAGA CGCA	GUUUUAGAGCUAGAAAU AGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACU UGAAAAAGUGGCACCGA GUCGGUGCUUUUUU	GTCTTGACGCAAGAGAC GCAGUUUUAGAGCUAG AAAUAGCAAGUUAAAA UAAGGCUAGUCCGUUA UCAACUUGAAAAAGUG GCACCGAGUCGGUGCU UUUUU
Rank 12	CGTCTT GACGC AAGAG ACGC	GUUUUAGAGCUAGAAAU AGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACU UGAAAAAGUGGCACCGA GUCGGUGCUUUUUU	CGTCTTGACGCAAGAGA CGCGUUUUAGAGCUAG AAAUAGCAAGUUAAAA UAAGGCUAGUCCGUUA UCAACUUGAAAAAGUG GCACCGAGUCGGUGCU UUUUU
