



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
**Facultad de Estudios Superiores Iztacala**



**Análisis de la ocupación del factor de transcripción MEOX2 en el  
Epigenoma del cáncer pulmonar: un análisis bioinformático  
comparativo.**

**Tesis**

**para optar por el grado de:  
LICENCIADA EN BIOLOGÍA**

**Presenta:**

**MARIANA GARCÍA JIMÉNEZ**

**Director:**

**DR. LEONEL ARMAS LÓPEZ**

**Comité asesor:**

**Dr. Federico Ávila Moreno**

**Dr. Felipe Vaca Paniagua**

**Dr. Carlos Alberto Peralta Álvarez**

**Dr. Luis Enrique Arias Romero**

**Los Reyes Iztacala, Estado de México, 2022**



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

## **AGRADECIMIENTOS**

A mis profesoras y profesores de la Facultad de Estudios Superiores Iztacala, por mi formación académica.

Al Dr. Leonel Armas López, por su conocimiento, apoyo y paciencia brindados.

A mis compañeras y compañero de laboratorio: Ana, Priscila y Octavio, por sus conocimientos y la experiencia compartida.

A mi comité asesor: Dr. Federico Ávila Moreno, Dr. Felipe Vaca Paniagua, Dr. Carlos Alberto Peralta Álvarez y Dr. Luis Enrique Arias Romero, por sus revisiones y correcciones a este trabajo.

## **DEDICATORIAS**

A la Dra. Elva Jiménez Hernández por ser una excelente madre y ejemplo, porque sin su invaluable apoyo no podría cumplir mis sueños.

A mi familia, por su apoyo económico, moral y afectivo.

A mis amigas, por sostenerme incontables veces.

A Eme, por ser clave en este camino.

A Honey, Eclipse y Zeus, por ser mis compañeras y compañero de vida.

A la Dra. Abigail Castro Contreras y la Psicóloga Nelly I. Juárez Pillado, por acompañarme y ayudarme en este proceso.

A las desobedientes por enseñarme tanto y acompañarme en la distancia.

A todas las mujeres y científicas que me anteceden, y a las que siguen luchando todos los días, porque sin ellas no podría estar aquí.

## GLOSARIO DE ABREVIATURAS

5mC	5-metilcitosina
AMPK	Proteína cinasa activada por AMP
BWA-MEM	Alineador Burrows-Wheeler
CDK	Cinasas dependientes de ciclina
ChIP-Seq	Secuenciación de la Inmunoprecipitación de cromatina
DNA	Ácido desoxirribonucleico
DNMTs	Metiltransferasas de DNA
EGFR	Receptor del factor del crecimiento epidérmico
EMT	Transición epitelio-mesénquima
EPOC	Enfermedad pulmonar obstructiva crónica
EZH2	Enhancer of Zeste Homologue 2
GEO	Gene expression omnibus
GSC	Células Troncales de Glioma
HDAC	Desacetilasa de Histonas
HDM	Desmetilasa de Histonas
HIF	Actor indicativo de hipoxia p53
HMT	Histona metiltransferasa
HOMER	Hypergeometric Optimization of Motif EnRichment
IARC	Agencia Internacional para la Investigación del Cáncer
IASLC	Asociación Internacional para el Estudio del Cáncer Pulmonar
IFN	Interferones
KRAS	Kirsten Rat Sarcoma Viral Oncogene
MACS	Model-based Analysis of ChIP-seq
MEC	Reorganización de la matriz extracelular
MEOX2	Mesenchyme Homebox-2
MMP	Metaloproteinasas de matriz
NCBI	Centro Nacional de Información Biotecnológica
NCI	Instituto Nacional del Cáncer
ncRNA	RNAs no codificantes
NSCLC	Carcinomas pulmonares de células no pequeñas
p53	Proteína inhibidora de cinasas dependiente de ciclina
PRC2	Complejo Represor Polycomb 2
pRB	Proteína de retinoblastoma
QC	Control de Calidad
SCLC	Cáncer de pulmón de células pequeñas
SRA	Archivo de lectura de secuencias
SRS	Radiocirugía estereotáctica
TMN	Tumor Nódulos Metástasis
TNF- $\alpha$	Factor de necrosis tumoral alfa
VEGF-A	Factor de crecimiento endotelial vascular A
WBRT	Radioterapia de cerebro completo
WHO	Organización Mundial de la Salud
GSEA	Análisis de enriquecimiento del conjunto de genes de prueba
ORA	Análisis de sobrerrepresentación
KEGG	Kyoto Encyclopedia of Genes and Genomes
RT-qPCR	PCR en tiempo real
cDNA	DNA complementario
mRNA	RNA mensajero
dNTPs	Dinucleotidos

# ÍNDICE

<b>1. INTRODUCCIÓN.....</b>	<b>9</b>
1.1. Definición de cáncer .....	10
1.2. Clasificación del cáncer .....	10
1.3. Origen y biología del cáncer .....	11
1.4. Señalización proliferativa.....	12
1.5. Evasión de los genes supresores de tumor .....	13
1.6. Resistencia a la muerte celular.....	14
1.7. Activación de la invasión y metástasis.....	15
1.8. Reprogramación epigenética.....	15
1.9. Cáncer de pulmón .....	17
1.10. Factores de riesgo del cáncer pulmonar.....	18
1.11. Clasificación histopatológica.....	19
1.12. Clasificación molecular .....	21
1.13. Sobrevida y diagnóstico .....	22
1.14. Mecanismos de regulación epigenética en cáncer pulmonar.....	23
1.16. Modificaciones de las histonas .....	26
1.17. Regulación y Factores de transcripción .....	27
<b>2. ANTECEDENTES.....</b>	<b>29</b>
2.1. Mesenchyme Homebox-2 (MEOX2) .....	30
2.2. MEOX2 en la actualidad en cáncer pulmonar.....	32
2.3. Regulación post-transcripcional de MEOX2 en cáncer pulmonar .....	33
2.4. Análisis del epigenoma en cáncer pulmonar por ensayos de CHIP-Seq y bioinformática.....	34
<b>3. JUSTIFICACIÓN.....</b>	<b>40</b>
<b>4. PLANTEAMIENTO DEL PROBLEMA .....</b>	<b>41</b>
<b>5. HIPÓTESIS.....</b>	<b>42</b>
<b>6. OBJETIVO GENERAL.....</b>	<b>42</b>
<b>7. OBJETIVOS PARTICULARES.....</b>	<b>42</b>
<b>8. MATERIALES Y MÉTODOS.....</b>	<b>43</b>
8.1. Fragmentación de la cromatina .....	42
8.2. Inmunoprecipitación de la cromatina .....	42
8.3. Construcción de bibliotecas y secuenciación masiva.....	43
8.4. Análisis bioinformático.....	43
8.5. Análisis de calidad.....	44
8.6. Alineamiento y análisis de calidad .....	44
8.7. Obtención de picos.....	44
8.8. Anotación genómica.....	45
8.9. Análisis de vías de señalización .....	45
8.10. Cultivo celular.....	45

8.11. RT-qPCR.....	46
8.12. Western blot .....	46
<b>9. ESTRATEGIA EXPERIMENTAL.....</b>	<b>49</b>
<b>10. RESULTADOS.....</b>	<b>50</b>
10.1. Preanálisis. Control de calidad de ChIP-Seq de MEOX2 y de secuencias control (input) provenientes de datos públicos. ....	49
10.2. Análisis central: Obtención del llamado de picos (Peak calling).....	52
10.3. Anotación genómica .....	54
10.4. Análisis bioinformático multi-comparativo .....	56
10.5. Análisis de vías de señalización .....	65
10.6. Validación experimental .....	69
<b>11. DISCUSIÓN.....</b>	<b>73</b>
<b>12. CONCLUSIONES.....</b>	<b>78</b>
<b>13. PERSPECTIVAS.....</b>	<b>79</b>
<b>14. BIBLIOGRAFÍA.....</b>	<b>80</b>

## ÍNDICE DE FIGURAS

<b>Figura 1.</b> Hallmarks del cáncer: La reprogramación epigenética como un “nuevo” mecanismo de modulación de procesos oncológicos.	<b>12</b>
<b>Figura 2.</b> Regulación epigenética.	<b>24</b>
<b>Figura 3.</b> Inmunoprecipitación de la cromatina seguida de secuenciación CHIP-Seq.	<b>35</b>
<b>Figura 4.</b> Proceso de secuenciación a través de plataforma Illumina.	<b>36</b>
<b>Figura 5.</b> Secuenciación masiva y análisis de la distribución de lecturas.	<b>37</b>
<b>Figura 6.</b> Diagrama experimental el cual indica flujo de trabajo para el desarrollo del análisis bioinformático del posicionamiento de MEOX2 sobre el epigenoma de cáncer pulmonar, así como el desarrollo de ensayos de validación experimental.	<b>48</b>
<b>Figura 7.</b> Control de calidad de secuencias sentido sin procesar de MEOX2.	<b>51</b>
<b>Figura 8.</b> Control de calidad de secuencias input 1 UW y 4 NY2.	<b>52</b>
<b>Figura 9.</b> Interacción del factor de transcripción MEOX2 a lo largo de todo el genoma.	<b>53</b>
<b>Figura 10.</b> Posicionamiento del factor transcripcional MEOX2 en el cromosoma 7	<b>54</b>
<b>Figura 11.</b> Posicionamiento de MEOX2 sobre el Epigenoma del cáncer pulmonar. La distribución-ocupación de MEOX2 ocurre principalmente sobre regiones génicas y distales intergénicas.	<b>55</b>
<b>Figura 12.</b> Predicción bioinformática de unión de MEOX2 sobre regiones genómicas principalmente distales al TSS.	<b>56</b>
<b>Figura 13.</b> Análisis diferencial del patrón de distribución promedio del llamado de picos de MEOX2 relativos al TSS, mediante el uso del algoritmo bioinformático MACS2 versus HOMER.	<b>57</b>
<b>Figura 14.</b> El análisis bioinformático comparativo del perfil promedio de los picos de MEOX2 sobre la región del TSS tratados con distintos DNA-inputs no muestra diferencias en el patrón de distribución relativa al TSS pero si al comparar el uso de los algoritmos HOMER versus MACS2.	<b>59</b>
<b>Figura 15.</b> Análisis comparativo del perfil promedio de los picos de MEOX2 en la región TSS.	<b>60</b>
<b>Figura 16.</b> Análisis comparativo del perfil promedio de los picos de ocupación de MEOX2 en la región TSS empleando distintos controles de DNA-input, y algoritmos empleados para el llamado de picos.	<b>61</b>
<b>Figura 17.</b> Análisis bioinformático comparativo que muestra la distribución de sitios de unión de MEOX2 con relación al sitio de inicio de la transcripción (TSS).	<b>62</b>
<b>Figura 18.</b> Análisis comparativo de la distribución de los sitios de unión de MEOX2 con relación al sitio de inicio de la transcripción (TSS).	<b>64</b>
<b>Figura 19.</b> Análisis bioinformático del enriquecimiento de vías funcionales de señalización bajo interacción por MEOX2.	<b>65</b>
<b>Figura 20.</b> Enriquecimiento funcional de MEOX2 mediante análisis de sobrerepresentación (ORA) mediante la plataforma de KEEG.	<b>66</b>
<b>Figura 21.</b> Volcano plot que representa el enriquecimiento funcional de MEOX2 mediante análisis de sobrerepresentación (ORA) por KEEG.	<b>67</b>
<b>Figura 22.</b> Enriquecimiento funcional de procesos moleculares probablemente bajo el control de MEOX2 mediante análisis de enriquecimiento del conjunto de genes (GSEA) a través de la plataforma KEEG.	<b>68</b>
<b>Figura 23.</b> Análisis de Volcano plot del enriquecimiento de vías de señalización funcional bajo el probable control de MEOX2 mediante análisis del enriquecimiento del conjunto de genes (GSEA) mediante plataforma KEEG.	<b>69</b>
<b>Figura 24.</b> Análisis de expresión de blancos génicos de MEOX2 en la línea celular A549 de cáncer de pulmón.	<b>70</b>
<b>Figura 25.</b> Análisis de expresión de proteína de MEOX2, GLI1 y EGFR en la línea celular A549.	<b>71</b>



## INDICE DE TABLAS

<b>Tabla 1.</b> Frecuencia de las anomalías moleculares presentes en SCLC y NSCLC.....	<b>22</b>
<b>Tabla 2.</b> Diseño de oligos para RT-qPCR.....	<b>46</b>
<b>Tabla 3.</b> Características de los datos de secuenciación de los 5 controles input y la muestra experimental .....	<b>50</b>

## RESUMEN

El factor de transcripción (TF) Mesenchyme Homebox-2 (MEOX2) se ha propuesto como un biomarcador tumoral involucrado en la resistencia al tratamiento fármaco-oncológica y pobre pronóstico clínico de pacientes con diagnóstico de carcinomas pulmonares de células no-pequeñas (NSCLC). Al respecto, se ha demostrado la sobreexpresión de MEOX2 en tejido tumoral, asociado con aberraciones del código de histonas. Por otro lado, no obstante ensayos de ChIP on chip han permitido identificar que MEOX2 regula la sobreexpresión de blancos genéticos, entre otros como la expresión a nivel de promotor del gen GLI-1 y se desconoce su impacto en la regulación del genoma y epigenoma en su conjunto en el cáncer pulmonar.

Con el objetivo de analizar y detectar a gran escala la ocupación de MEOX2 sobre el epigenoma del cáncer pulmonar, se emplearon bases de datos generadas por nuestro grupo de investigación a través de ensayos de ChIP-Seq sobre células de cáncer pulmonar A549. A partir de ello, se llevaron a cabo ensayos de análisis bioinformático multi-comparativo, con el objetivo adicional de demostrar que el uso de datos depositados en repositorios públicos de ChIP-Seq como el consorcio ENCODE constituyen datos de suficiente calidad para llevar a cabo análisis multi-comparativos de secuenciación a nivel internacional; adicional a ello, nuestros resultados *in silico* fueron validados a través de ensayos de RT-qPCR y Western Blot.

Nuestros resultados a nivel bioinformático sugieren la interacción de MEOX2 sobre regiones distales del genoma de 10 a 100 kb (50%), seguido de regiones mayores a 100 kb (40%) en relación al TSS tanto downstream como upstream.; así mismo identificamos su posicionamiento sobre genes de importantes vías de señalización probablemente involucradas en el desarrollo neoplásico pulmonar, como el caso de la vía Sonic Hedgehog.

Los resultados del análisis bioinformáticos indicaron que el TF MEOX2 se posiciona a lo largo de todo el genoma, incluso más allá de secuencias

promotoras de 0 a 3 kb, sugiriendo su participación en la regulación amplia del epigenoma involucrada en procesos oncológicos pulmonares.

Por último nuestro análisis multi-comparativo evidenció que el uso de controles del DNA input de células A549, para el procesamiento de resultados de ChIP-Seq, aseguran el enriquecimiento reproducible del perfil de secuencias “picos” en el epigenoma de células A549 de cáncer de pulmón, permitiendo asegurar que el uso de DNA Input de bases de datos públicos son confiables para el desarrollo de análisis computacional masivo del epigenoma a través de ensayos de ChIP-Seq.

## **ABSTRACT**

The transcription factor (TF) Mesenchyme Homebox-2 (MEOX2) has been proposed as a tumor biomarker involved in the resistance to drug-oncological treatment and poor clinical prognosis of patients diagnosed with non-small cell lung carcinoma (NSCLC). In this regard, the overexpression of MEOX2 in tumor tissue, associated with aberrations of the histone code, has been demonstrated. On the other hand, despite ChIP on-chip assays, it has been possible to identify that MEOX2 regulates the overexpression of genetic targets, among others, such as expression at the GLI-1 promoter level, its impact on the regulation of the genome and epigenome is still unknown in lung cancer diagnosis.

In order to analyze and detect the occupation of MEOX2 on the epigenome of lung cancer on a large scale, databases generated by our research group through ChIP-Seq assays on A549 lung cancer cells were used. Based on this, multi-comparative bioinformatic analysis tests were carried out, with the additional objective of demonstrating that the use of data deposited in public ChIP-Seq repositories as well as in ENCODE, constitute data of sufficient quality to carry out multi comparative sequencing analysis at an international level; additionally, our in silico results were validated through RT-qPCR and Western Blot assays.

Our results at the bioinformatic level demonstrated the interaction of MEOX2 on distal regions of the genome from 10 to 100 kb (50%), followed by regions greater than 100 kb (40%) in relation to the TSS, both downstream and upstream.; Likewise, we identified its position on genes of important signaling pathways probably involved in lung neoplastic development, as in the case of the Sonic Hedgehog pathway.

The results of bioinformatic analysis indicated that the transcription factor MEOX2 is positioned throughout the entire genome, even beyond promoter sequences of 0 to 3 kb, suggesting its participation in the broad regulation of the epigenome involved in pulmonary oncological processes.

Finally, our multi-comparative analysis showed that the use of A549 cell input DNA controls for processing ChIP-Seq results ensures the duplication of "peak" enrichment of the sequence profile in the cell epigenome A549 for breast cancer, assuring that the use of DNA Input from public databases are reliable for the development of massive computational analysis of the epigenome through ChIP-Seq assays.

# 1. INTRODUCCIÓN

## 1.1. Definición de cáncer

El cáncer es definido por el Instituto Nacional del Cáncer (NCI por sus siglas en inglés) como una enfermedad en la cual algunas células crecen sin control y pueden diseminarse a diferentes partes del cuerpo (Instituto Nacional del Cáncer, 2011). La Organización Mundial de la Salud (WHO por sus siglas en inglés) lo define como un amplio grupo de enfermedades que pueden afectar a cualquier parte del organismo, cuya característica definitoria es la multiplicación de células anormales que se extienden más allá de sus límites habituales e invaden partes adyacentes del cuerpo o se propagan por otros órganos (Ferlay J. *et al.*, 2020). Por último, una definición más amplia del cáncer que toma en cuenta características moleculares, bioquímicas y celulares, lo define como un conjunto de factores genéticos, epigenéticos y estructurales que conducen a la transformación progresiva fenotípica y estructural de las células en tejidos, epitelios y órganos (Hanahan & Weinberg, 2000)

## 1.2. Clasificación del cáncer

Existen diversas clasificaciones del cáncer basadas en criterios como la etiológica, la localización, histología y genética las cuales son utilizadas como determinantes primarios en el diagnóstico y tratamiento. De acuerdo con el tipo de tejido se clasifican en:

**Carcinoma:** Se denomina carcinoma cuando se origina en las células epiteliales que recubren la piel y las mucosas. Dentro de este grupo se encuentra el adenocarcinoma y el carcinoma de células escamosas o carcinoma epidermoide.

**Sarcoma:** Este tipo se origina en los huesos (osteosarcoma) y en los tejidos blandos, como en el músculo, cartílago, tejido adiposo, vasos sanguíneos, y vasos linfáticos, también se origina en tejidos fibrosos como tendones y ligamentos. (He *et al.*, 2012)

**Mieloma múltiple:** Este se origina en las células plasmáticas de la médula ósea.

**Leucemia:** En este subtipo no se forman tumores sólidos y engloba aquellos que se originan en la médula ósea. Existen varios tipos de leucemias que se clasifican de acuerdo con la rapidez de propagación y el tipo celular involucrado, tales como, leucemia mielocítica aguda (LMA), leucemia mielocítica crónica (LMC) Leucemia linfocítica o linfoblástica aguda (LLA), leucemia linfocítica, linfocítica o linfoblástica crónica (LLC) y policitemia Vera o eritremia.

**Linfoma:** Se denomina linfoma a aquel que se origina específicamente en los linfocitos: células T o células B sin embargo a diferencia de las leucemias, si se presentan tumores sólidos o linfomas extranodales que afectan a los ganglios linfáticos en órganos y sitios concretos en el cuerpo (DeVita *et al.*, 2017).

### 1.3. Origen y biología del cáncer

Douglas Hanahan y Robert A. Weinberg (Hanahan & Weinberg, 2000, 2011) describieron las capacidades biológicas distintivas del cáncer, las cuales incluyen:

- a) Mantenimiento de la señalización proliferativa
- b) Evasión de genes supresores del crecimiento
- c) Resistencia a la muerte celular
- d) Inmortalidad replicativa
- e) Inducción de angiogénesis
- f) Activación de invasión y metástasis
- g) Reprogramación del metabolismo energético
- h) Evasión de la respuesta inmune
- i) Inestabilidad del genoma y mutación
- j) Inflamación tumoral

Además de las neoplásicas, los tumores contienen un repertorio de células reclutadas, aparentemente normales, que contribuyen a la adquisición de rasgos distintivos mediante la creación del “microambiente tumoral” (Hanahan & Weinberg, 2011). Al respecto, en 2022 Hanahan agregó 4 sellos distintivos emergentes que consideran el desbloqueo de la plasticidad fenotípica, la

reprogramación epigenética no mutacional, los microbiomas polimórficos y las células senescentes (Hanahan, 2022).



**Figura1. Hallmarks del cáncer: La reprogramación epigenética como un “nuevo” mecanismo de modulación de procesos oncológicos.** A pesar de las evidencias sobre la reprogramación epigenética en la modulación de la expresión génica en general, y en los procesos oncológicos en particular; es hasta 2022 que Hanahan agrega la reprogramación epigenética no mutacional como una capacidad del cáncer que podría interactuar con cualquiera de las características (tomada y modificada de Hanahan, 2022).

Con base en todo lo anterior, a continuación se mencionan los principales *Hallmarks* del cáncer que guardan relación con la hipótesis que se aborda en el presente trabajo experimental de tesis.

#### 1.4. Señalización proliferativa

La proliferación inapropiada de las células en el desarrollo del cáncer resulta en parte de la activación anormal de los mecanismos regulatorios que usualmente ocurren durante el desarrollo embrionario, el crecimiento fisiológico y el mantenimiento homeostático de los tejidos en todo el cuerpo. En el contexto del



cáncer, esta característica engloba un conjunto complejo de señales inductivas que propician el inicio y la progresión del ciclo de producción de células hijas (Bast, 2016).

Los mecanismos de señalización involucran alteraciones mutacionales de genes definidos como oncogenes, los cuales codifican proteínas alteradas en su estructura o en su función y se encuentran en niveles más abundantes que en condiciones normales (Brandt-Rauf, 1992).

Algunos oncogenes que mantienen la señalización proliferativa en cánceres humanos incluyen al receptor del factor del crecimiento epidérmico (*EGFR*) y a los transductores de señal en la vía descendente KRAS-RAF-MEK-ERK, que a su vez procesan y transmiten señales estimulantes del crecimiento a través de la fosforilación de proteínas de la maquinaria de la división celular (Hanahan & Weinberg, 2011; Martin, 2003)

Otros mecanismos que activan el crecimiento tumoral son: la sobreexpresión de receptores, que vuelve a las células neoplásicas hipersensibles a los factores de crecimiento; y las mutaciones o deleciones específicas, que provocan una señalización independiente del ligando (Witsch *et al.*, 2010).

### **1.5. Evasión de los genes supresores de tumor**

Además de mantener las señales que promueven el crecimiento tumoral, las células neoplásicas adquieren la capacidad de evadir mecanismos de regulación del ciclo celular, debido a mutaciones de pérdida de función en los genes supresores de tumor, los cuales median diferentes puntos de control de la división celular en primera instancia a través de la proteína retinoblastoma (pRB), así como a través de inhibidores de cinasas dependientes de ciclina (CDK) (Evan & Vousden, 2001).

Las proteínas de la familia RB constituyen una vía de control en la progresión de la fase G1 por medio de la inhibición de los factores de transcripción E2F, estas proteínas se disocian de E2F y se activan al ocurrir su fosforilación mediante

cinastas dependientes de ciclina (Dick & Rubin, 2013). La desregulación de esta vía incluye la delección del gen *RB* o desregulación de las CDK por sobre-activación directa o pérdida genética de sus inhibidores. En el cáncer de pulmón, por ejemplo, la pérdida de *RB* predomina en tumores de células pequeñas; mientras que la expresión de p16 ocurre en la mayoría de los carcinomas NSCLC (Sherr & McCormick, 2002).

Por otro lado, p53 es una proteína codificada por el gen *TP53* que juega un papel importante en el arresto del ciclo celular, la apoptosis y en los mecanismos de reparación del DNA (Chumakov, 2000). Las mutaciones de p53 inducen mecanismos adaptativos ante el estrés oxidativo y proteo-tóxico para generar una respuesta (Mantovani *et al.*, 2019).

## **1.6. Resistencia a la muerte celular**

Los mecanismos de apoptosis representan al menos 2 vías de muerte celular conservada evolutivamente responsables de la eliminación controlada de células durante el desarrollo y el mantenimiento de la homeostasis del organismo. En este sentido los defectos en las vías inductoras de apoptosis pueden conducir a la expansión de una población de células neoplásicas (Hanahan & Weinberg, 2017).

Existen dos vías de inducción de apoptosis: *i)* la extrínseca activada por el acoplamiento del ligando de los receptores de muerte en la superficie celular, y *ii)* la intrínseca o mitocondrial mediada por miembros de la familia BCL-2. En ambas vías, proteasas específicas de cisteína (caspasas) son activadas y escinden los sustratos celulares, produciendo cambios bioquímicos y morfológicos en las células en proceso de muerte (Igney & Krammer, 2002).

Los mecanismos de evasión de la apoptosis ayudan en la transformación oncogénica en múltiples etapas, al facilitar el crecimiento tumoral sostenido, la supervivencia durante el proceso metastásico y la resistencia a la terapia, como se ha observado en la expresión desregulada de varios miembros de la familia Bcl-2 en tumores malignos de distinto origen tisular (D'Aguanno & Del Bufalo, 2020).

### **1.7. Activación de la invasión y metástasis**

La metástasis es la diseminación de las células neoplásicas desde el tumor primario a los tejidos circundantes y órganos distales (Seyfried & Huysentruyt, 2013). Las lesiones metastásicas masivas son la principal causa de morbilidad y mortalidad por cáncer, puesto que conducen a la insuficiencia orgánica grave y, por lo tanto, a la muerte del paciente (Spano *et al.*, 2012).

Por otro lado, la invasión es la capacidad de las células tumorales para romper la membrana basal y penetrar el estroma subyacente. Durante la invasión, se regulan tres procesos importantes: la adhesión, la reorganización de la matriz extracelular (MEC) y la motilidad. Aunque para la metástasis la invasión es necesaria, la capacidad de invadir no es suficiente (Welch & Hurst, 2019); para que ocurra las células neoplásicas deben desprenderse del tumor primario, penetrar en los sistemas circulatorio y linfático, evadir el ataque inmunológico, extravasarse e invadir y proliferar en órganos distantes (van Zijl *et al.*, 2011).

Un evento clave en la promoción de la migración e invasión de células tumorales es la activación del programa de la EMT. A través de factores de transcripción, como las proteínas con motivos de dedos de zinc de la familia SNAIL (SNAIL1 y SNAIL2); las proteínas con motivo de hélice-bucle-hélice de la familia TWIST (TWIST1 y TWIST2); y las proteínas Homeobox de unión E-box de dedos de zinc (ZEB1 y ZEB2), se inhiben la expresión de genes y se codifican proteínas de unión epitelial, desensamblando uniones adherentes, desmosomas, uniones estrechas y pérdida de polaridad apical-basal (Yeung & Yang, 2017). También se han identificado otros dos modos de invasión: la migración colectiva a través de láminas epiteliales, o individual mediante células ameboides (van Zijl *et al.*, 2011).

### **1.8. Reprogramación epigenética**

Los mecanismos de regulación epigenética se refieren a la detección cuantificable de cambios bioquímicos heredables y estables en la expresión génica, libres de alteraciones en la secuencia primaria del DNA. En las células humanas, las marcas epigenéticas se establecen una vez que se han diferenciado o al término

de cada ciclo celular. Sin embargo, durante el desarrollo neoplásico, algunas células se someten a una "reprogramación" epigenética, que implica la adquisición *de novo* de programas de desarrollo, a través del establecimiento de un conjunto diferente de marcas lo que altera los patrones de expresión génica y de identidad celular (Morgan *et al.*, 2005).

La evidencia acumulada sugiere que los mecanismos de regulación epigenética, tales como la metilación del DNA, la modificación de histonas, la alteración por remodeladores de cromatina dependientes de ATP, los RNAs no codificantes, así como la distribución temporal espacial de la cromatina en el núcleo, tienen una función central en los eventos iniciadores, y de heterogeneidad celular en el establecimiento y progresión del cáncer (Sharma *et al.*, 2010).

La metilación del DNA es una modificación que ha funcionado como marcador epigenético y se ha relacionado con la activación de elementos transponibles y protooncogenes mediante hipometilación del DNA. Además, la hipometilación global contribuye a la pérdida de impronta, como se ha demostrado para IGF2/H19 que promueve la sobreexpresión del gen *IGF2*, siendo una mutación iniciadora de la carcinogénesis. Por otro lado, la hipermetilación se asocia con el silenciamiento de los genes supresores de tumores, como *RAS*, que es un mediador clave de la señalización mitogénica y la proliferación celular, *BRCA1*, que codifica proteínas que reparan daños en el DNA y *BCL2*, el cual es un gen clave que participa en la vía intrínseca de apoptosis celular (Miranda Furtado *et al.*, 2019).

Una de las características del cáncer es la desregulación del metabolismo, el cual tiene una dinámica sinérgica con las modificaciones epigenéticas, ya que a través de esta se regula la expresión de muchos genes metabólicos afectando la biosíntesis de macromoléculas y la producción de energía. Por ejemplo el aumento de acetilación de SDHA (complejo de succinato deshidrogenasa, subunidad A) mediada por cMYC activa la degradación de SIRT3 desacetilasa, esto lleva a la desactivación de SDHA y como consecuencia hay una acumulación de succinato. Este aumento de succinato inhibe la actividad de las histonas desmetilasas, desencadenando la trimetilación de la histona 3 lisina 4 y la

expresión de genes específicos del tumor que deriva en su progresión (Li *et al.*, 2020).

Adicional a ello se ha demostrado heterogeneidad intratumoral que comprende varias subpoblaciones de células con distintos fenotipos y propiedades biológicas generando células con mayor capacidad de expansión e invasión, como la histona enlazadora H1.0, que se expresa y reprime dinámicamente en subpoblaciones de células neoplásicas dentro de varios tipos de tumores (Hanahan, 2022).

### **1.9. Cáncer de pulmón**

El cáncer es una de las principales causas de muerte en el mundo. El crecimiento y envejecimiento poblacional está asociado con el aumento en su incidencia y mortalidad; por otro lado, la prevalencia y distribución de sus principales factores de riesgo están asociados con el desarrollo socioeconómico (Sung *et al.*, 2021).

En 2020 se estimaron 19,3 millones de casos nuevos y 10 millones de muertes por cáncer, de los cuales se distingue que el de mama femenino es el más comúnmente diagnosticado (11,7% del total de casos), seguido por el cáncer de pulmón (11,4%), el colorrectal (10,0%), de próstata (7,3%) y de estómago (5,6%) (Sung *et al.*, 2021).

El de pulmón, conformado por tumores malignos broncogénicos del epitelio de las vías respiratorias, es la principal causa de muerte por cáncer (18,0%). Dicha neoplasia rara vez aparece antes de los 40 años y las cifras aumentan hasta los 80; la probabilidad de que surja cáncer de pulmón calculada para toda la vida es de 8% en varones y 6% en mujeres, y su incidencia varía geográficamente entre grupos raciales y étnicos (Schabath & Cote, 2019).

En México, en 2020 se registraron 7 mil 588 casos nuevos y 6 mil 733 muertes por cáncer de pulmón. Esto representa un problema de salud pública dado que la mayoría de los casos son diagnosticados en etapas avanzadas y la supervivencia a largo plazo es poco satisfactoria pese a los avances en su conocimiento (*Cancer*

Today, 2020).

### **1.10. Factores de riesgo del cáncer pulmonar.**

Anteriormente, más del 90% de los casos de cáncer de pulmón se atribuían al consumo de tabaco, sin embargo, en las últimas décadas se ha evidenciado que una parte de la incidencia ocurre en personas no fumadoras, lo que ha llevado a considerar una etiología multifactorial que considera carcinógenos ambientales, ocupacionales, factores genéticos y epigenéticos que regulan la expresión o represión de genes involucrados en el desarrollo tumoral (Malhotra *et al.*, 2016).

El consumo de tabaco es el principal factor de riesgo para el desarrollo de cáncer pulmonar; el riesgo aumenta según el número de cigarrillos consumidos por día y la edad de inicio del consumo; igualmente, la exposición involuntaria en el hogar durante la niñez, o en el trabajo, también incrementa riesgo de adquirirlo (Brenner *et al.*, 2011).

El cigarrillo contiene diversas sustancias, algunas de ellas han sido evaluadas y determinadas como carcinógenos por la Agencia Internacional para la Investigación del Cáncer (IARC por sus siglas en inglés). Los carcinógenos se transforman enzimáticamente a medida que el organismo intenta convertirlos en productos más fáciles de excretar, conduciendo a su activación metabólica formando aductos de DNA (Guengerich, 2006).

Si estos no son reparados eficazmente generan mutaciones en el gen supresor de tumores *TP53* y el oncogén *KRAS* (Kirsten rat sarcoma viral oncogene) (Hecht, 1999). Así mismo, los iones de níquel presentes en el humo de tabaco inducen desacetilación de las histonas H2A, H2B, H3 y H4 e incrementos en la marca H3K9me2, asociada a represión transcripcional (T. Rauch *et al.*, 2007).

La contaminación ambiental también es considerada carcinógeno pulmonar por la IARC, aunque su estudio es limitado por falta de indicadores directos para su correlación, empero, las mediciones ambientales de partículas finas con

frecuencia superan las pautas nacionales e internacionales y sugieren un aumento en el riesgo (Loomis *et al.*, 2014).

Las exposiciones ocupacionales tienen un papel importante en el desarrollo de cáncer pulmonar, el riesgo aumenta entre las y los trabajadores expuestos al amianto, sílice, radón, metales pesados, hidrocarburos aromáticos policíclicos, humo de leña y cocina entre otros agentes (Malhotra *et al.*, 2016).

Debido a que solo una pequeña cantidad de fumadores desarrollan cáncer se considera su susceptibilidad genética. Se han identificado relaciones con la variación del polimorfismo de un solo nucleótido en diferentes *loci*, entre ellos 15q25 (Herbst *et al.*, 2008) con dos genes que codifican subunidades del receptor alfa nicotínico de acetilcolina regulado por la exposición a la nicotina (Amos *et al.*, 2008). Sin embargo, aunque las mutaciones y las alteraciones del número de copias desempeñan un papel en la oncogénesis, las alteraciones epigenéticas son de hecho más frecuentes que las mutaciones somáticas en el cáncer de pulmón (Brzezińska *et al.*, 2013).

Se ha visto que el padecimiento de afecciones pulmonares previas, como EPOC (Enfermedad Pulmonar Obstructiva Crónica), neumonía y tuberculosis son fuentes de desarrollo de la neoplasia pulmonar sobre todo en pacientes no fumadores (Brenner *et al.*, 2011), esto debido a procesos inflamatorios prolongados que conducen a daño tisular, fibrosis, formación de cicatrices y daño genómico (Wong *et al.*, 2020). Por último, existe evidencia de la influencia de los factores dietéticos en la aparición del cáncer pulmonar; el colesterol es esencial para el funcionamiento del metabolismo celular, sin embargo, su sobrecarga produce respuestas citotóxicas e inflamatorias tras la activación del sistema inmunológico (Gowdy & Fessler, 2014).

### **1.11. Clasificación histopatológica**

La clasificación histopatológica del cáncer de pulmón intenta correlacionar la morfología histológica de las células tumorales con sus características biológicas

en respuesta a las alteraciones moleculares. Dicha clasificación se encuentra en congruencia como herramientas de diagnóstico, así como en la decisión de los esquemas de tratamiento y su evaluación clínica a nivel pronóstico, por lo que, se ha clasificado en dos grandes grupos: **i)** cáncer de pulmón de células pequeñas (SCLC por sus siglas en inglés) el cual comprende cerca del 15% de los casos, exhibiendo diferenciación morfo-histológica de tipo neuroendocrina, mientras que **ii)** el cáncer de pulmón de células no-pequeñas (NSCLC por sus siglas en inglés) el cual engloba cerca del 85% de los casos en todo el mundo (Parikh, 2019), este se subdivide en tres principales tipos histopatológicos, como a continuación se indica:

**Carcinomas de tipo escamoso:** Comprende el 16% de los casos del grupo NSCLC, localizados en la región central del pulmón a lo largo de las principales vías respiratorias. Las células presentan queratinización y puentes intercelulares, exhibiendo un patrón de crecimiento de anidado sólido.

**Carcinomas de células grandes:** Su incidencia a nivel mundial es del 32%, se presenta en la periferia pulmonar, preferentemente exhibe un aspecto necrótico. Las células tumorales presentes son grandes y de forma poligonal con núcleos pleomórficos y vesiculares, así mismo conforma láminas sólidas libre de patrones irregulares.

**Adenocarcinomas:** Se presenta en el 42% de los casos de NSCLC, siendo el subtipo más común, el cual conforma una masa localizada en la periferia con fibrosis central y pliegues pleurales, su origen es epitelial moco-productor con diferenciación granular y secretor de mucinas (Zheng, 2016).

Esta agrupación histopatológica ha sido actualizada y reclasificada con nuevas características histológicas en función de nuevos marcadores inmunohistoquímicos, y características moleculares que permiten direccionar la aplicación de terapias dirigidas (Parikh, 2019), por lo que dentro de este grupo se subclasifican entre otros en:

**Adenocarcinoma lepidico:** Este denota células tumorales que se extienden a



los largo de estructuras alveolares preexistentes con engrosamiento esclerótico de los tabiques alveolares, además se caracterizan por presencia de estroma desmoplásico, propagación por espacios aéreos, invasión linfática o vascular, invasión pleural y necrosis tumoral.

**Adenocarcinoma acinar:** Presenta células tumorales dispuestas en una estructura glandular sobre un estroma sobre un estroma fibroelástico.

**Adenocarcinoma papilar:** Las células tumorales forman una arquitectura gpapilar con células tumorales que recubren la superficie de núcleos fibrovasculares ramificados, esta última característica representa la diferencia con el adenocarcino micropapilar.

**Adenocarcinoma micropapilar:** En este subtipo las células aparecen en grupos pequeños y sólidos separados y se asocia con pobre pronóstico desde la etapa inicial en comparación con otros adenocarcinomas.

**Adenocarcinoma sólido:** Las células tumorales forman láminas y carecen de cualquier otro patrón reconocible, presenta marcadores comúnmente asociados con la diferenciación de células escamosas como p63 y p40 (Travis *et al.*, 2013).

## 1.12. Clasificación molecular

Los análisis genéticos han demostrado diferentes alteraciones a nivel molecular en los diferentes tipos de cáncer pulmonar: anomalías en oncogenes y genes supresores de tumor son utilizados como marcadores predictivos y/o blancos terapéuticos. En general, los principales promotores oncogénicos mejor conocidos son las mutaciones y reordenamientos de *EGFR*, *KRAS* y *ALK* en los tumores tipo adenocarcinoma, en los que el oncogén *KRAS* se encuentra mutado en aproximadamente el 30% de los pacientes. Mientras que los genes *MYC*, *Cyclina D1* y *EGFR* se encuentran amplificados promoviendo la sobreexpresión en 2.5-10%, 5% y 6%, respectivamente; Así mismo, la sobreexpresión de *BCL2* está

involucrada en aproximadamente el 25% de los casos. Otras mutaciones importantes identificadas recientemente involucran a los genes *BRAF*, *HER2* y *PIK3CA*, así como translocaciones en los genes *ROS1* y *RET* (Calvayrac *et al.*, 2017).

**Tabla 1.** Frecuencia de las anomalías moleculares presentes en SCLC y NSCLC (Tomado y modificado de (DeVita *et al.*, 2017).

<b>Aberraciones</b>	<b>SCLC</b>	<b>NSCLC</b>
Inestabilidades de microsatélites	~35%	~22%
Bucles autocrinos	Receptor de GRP/GRP; SCF/KIT	TGF- $\alpha$ /EGFR heregulina/ HER2/neu; HGF/MET
Mutación puntual <i>RAS</i>	<1%	15%-20%
Mutación de <i>EGFR</i>	<1%	10-40%
<i>EMLK4-ALK</i>	0%	3%-7%
<i>ROS1</i>	0%	1%
<i>RET</i>	0%	1%
<i>ERBB2</i>	0%	2%-4%
<i>PIK3CA</i>	0%	1%-3%
<i>FGFR</i>	0%	20%
Sobreexpresión de la familia MYC	15%-30%	5%-30%
Inactivación de <i>TP53</i>	~90%	~50%
Inactivación de <i>RB</i>	~90%	15%-30%
Inactivación <i>p16<sup>INKA4</sup></i>	0%-10%	30%-70%
Inactivación <i>LKB</i>	~40%-60% (IHC)	20%-40%
Pérdida alélica frecuente	3p,4p,4q,5q,8p,10q,13q,17p,22q	3p,6q,8p,9p,13p,17p,19q
Actividad de la telomerasa	~100%	80%-85%
Expresión de <i>BCL2</i>	75%-95%	10%-35%
GRB: péptido liberador de gastrina; SCF: Factor de células troncales; TGF- $\alpha$ : Factor $\alpha$ de crecimiento transformante		

### 1.13. Sobrevida y diagnóstico

Los estadios clínicos del cáncer de pulmón se encuentran relacionados con la etapa de evolución clínica al momento del diagnóstico; dichas etapas clínicas arrojan una tasa de sobre-vida a cinco años del 56% para el estadio clínico I,

hasta una tasa de supervivencia a 5 años de solo el 4% en el estadio clínico IV (Woodard *et al.*, 2016).

La evaluación diagnóstica de pacientes con sospecha de cáncer de pulmón incluye el diagnóstico de tejidos (Biopsia presuntiva y/o post-operatorio) con una evaluación clínica-oncológica completa de la estadificación. Adicionalmente, el diagnóstico histológico puede obtenerse a través de citología de esputo, toracocentesis, biopsia de ganglio linfático accesible, broncoscopia, aspiración con aguja transtorácica, toracoscopia o toracotomía. También es posible realizar una evaluación inicial de la enfermedad metastásica basada en la historia clínica y el examen físico del paciente, pruebas de laboratorio, tomografía computarizada de tórax, tomografía por emisión de positrones y confirmación tisular de la afectación mediastínica (Collins, 2007).

#### **1.14. Mecanismos de regulación epigenética en cáncer pulmonar**

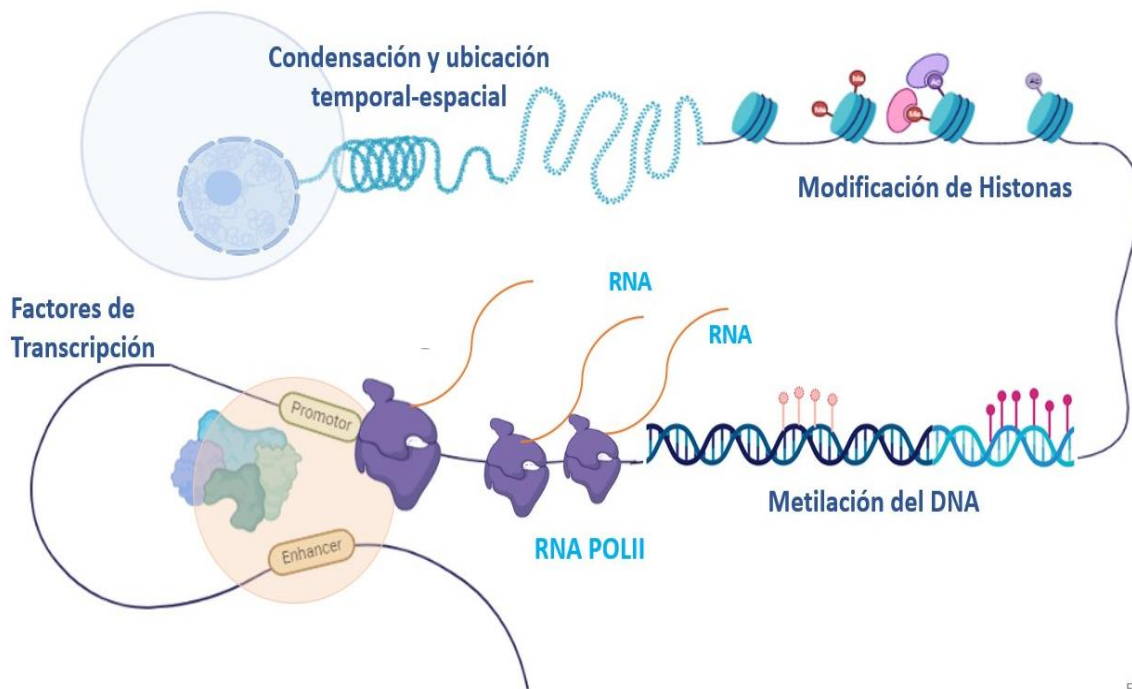
Como ya se mencionó anteriormente, los mecanismos epigenéticos, mediados por modificaciones bioquímicas propias del DNA o mediante modificaciones en proteínas, modulan la estructura de la cromatina y confieren un programa de expresión génica diferencial sin cambios en la secuencia primaria de DNA, juegan un papel importante en la modulación de la expresión genética en células normales y neoplásicas. (Nebbio *et al.*, 2018).

Con base en lo anterior, los mecanismos de regulación transcripcional sobre un locus genético está determinado en gran medida por la conformación de la cromatina, dicha región genética debe ser accesible a los factores reguladores, así como a la maquinaria transcripcional. La cromatina está determinada por la estrecha interacción entre los represores de la familia Polycomb, como los activadores de la familia Trithorax, además de los complejos remodeladores de la cromatina (Wadowska *et al.*, 2020).

Los distintos niveles de compactación del DNA permiten el control de la accesibilidad a la cromatina para la replicación, reparación y recombinación del

DNA. En cáncer de pulmón se han detectado incrementos en las marca de histona H3K9me2, así como desacetilación de las histonas H2A H2B H3 y H4, provocando represión génica, en parte debido al exceso de iones de níquel presentes en el humo de tabaco.

En el cáncer de pulmón se ha observado hipometilación global de oncogenes (*KRAS*,*HRAS*) e hipermetilación en genes supresores de tumor (*p16/CDKN2A*, *MGMT* y *CDH13*). Estos mecanismos alterados, a su vez, modifican la expresión de los genes en la cual se requieren distintos elementos reguladores que posibilitan o reprimen la transcripción, tales como los factores de transcripción, o bien las proteínas de unión al DNA que reclutan distintos elementos o median la interacción temporal-espacial de la transcripción (Cuyàs *et al.*, 2021).



**Figura 2. Regulación epigenética.** Se muestra de forma esquemática los distintos mecanismos de regulación epigenética que promueven la expresión o represión génica (Creada en Biorender).

### 1.15. Metilación del DNA

La metilación del DNA en los dinucleótidos CpG súper-puestos en secuencias promotoras se asocia comúnmente con el silenciamiento de la expresión genética, por impedimento estérico, lo cual permite explicar su importancia en la regulación de la expresión genética. La metilación del DNA es catalizada por una familia de metil-transferasas del DNA (DNMTs), las cuales transfieren un grupo metilo de S-adenil metionina al quinto carbono de un residuo de citosina para conformar 5-metilcitosina (5mC). Existen DNMTs *de novo* (DNMT3a y DNMT3b) que pueden establecer patrones nuevos de metilación sobre el DNA no modificado, así como DNMTs de mantenimiento como la DNMT1 la cual funciona durante la replicación del DNA capaz el patrón de metilación de la hebra del DNA parental (Moore *et al.*, 2013). El aumento de la actividad de las DNMTs es uno de los factores asociados con el aumento de la susceptibilidad a la hipermetilación de islas CpG en genes supresores de tumor en distintos tipos de cáncer en humanos, las cuales se encuentran reguladas positivamente en el NSCLC (H. Kim *et al.*, 2006). Por otro lado las proteínas de la familia TET eliminan las marcas de metilación a través de mecanismos indirectos, y se ha observado reducción de su actividad provocando hipermetilación de genes supresores de tumor, esto debido a una característica común del ambiente tumoral que es la hipoxia (Thienpont *et al.*, 2016).

Un evento temprano en el cáncer de pulmón es la inactivación de genes supresores de tumor, a través de la hipermetilación de secuencias promotoras. Algunos de los más estudiados son *p16 INK4a*, *RASSF1A*, *APC*, *RARβ*, *CDH1*, *CDH13*, *DAPK*, *FHIT*, *SOX2* y *MGMT* (Langevin *et al.*, 2015) que participan en diferentes procesos como la regulación del ciclo celular, la diferenciación, la reparación del DNA y el metabolismo energético. La metilación en islas CpG de genes asociados a factores de transcripción tipo Homeobox, representan un evento común en el cáncer de pulmón en estadios clínicos I; mientras que por el contrario, en estadios avanzados, otros genes comúnmente hipermetilados, son *hDAB2IP*, miembro de la familia de activadores de Ras GTPasa, *H-Cadherin* cuya proteína se relaciona con la superfamilia de cadherinas implicadas en la adhesión

celular, *DAL-1* que codifica proteínas estructurales del citoesqueleto y *FBN2* componente de las microfibrillas del tejido conjuntivo (Kikuchi *et al.*, 2005).

En NSCLC la hipometilación generalizada también se ha relacionado con la pérdida de impronta y la inestabilidad genómica, esta última se refiere a una mayor susceptibilidad de presentar y mantener mutaciones en el DNA, lo cual puede resultar en la activación de oncogenes debido a que los cambios en la secuencia de nucleótidos, asociados a sustituciones, deleciones, inserciones, translocaciones o amplificaciones derivan en un efecto deletéreo de los subsecuentes productos de la expresión génica como las proteínas; esta tiende a ocurrir en elementos nucleares, elementos LTR, duplicados segmentarios y regiones subteloméricas (T. A. Rauch *et al.*, 2008).

#### **1.16. Modificaciones de las histonas**

En los eucariontes el DNA se encuentra organizado formando complejos denominados nucleosomas, a su vez formados por un octámero de proteínas denominadas histonas. Dichas histonas son blanco de modificaciones postraduccionales, tales como metilación, acetilación, fosforilación, ubiquitinación entre otras, que afectan la accesibilidad de la maquinaria transcripcional, al modular los estados de heterocromatina y eucromatina. Estos cambios se llevan a cabo por complejos enzimáticos modificadores, cuyas mutaciones han sido asociadas con el inicio y la progresión del cáncer (Füllgrabe *et al.*, 2011). Las enzimas histonas acetiltransferasas (HAT), inducen la descondensación de la cromatina al acetilar las colas de histonas; por el contrario, las enzimas histonas desacetilasas (HDAC) eliminan los grupos acetilo y promueven la unión de las histonas al DNA provocando un estado heterocromático transcripcionalmente inactivo. Las histonas metiltransferasas (HMT) inhiben la transcripción según el residuo de histona que modifiquen y por último las desmetilasas (HDM) contrarrestan las modificaciones de las HMT (Allis *et al.*, 2007). Algunos ejemplos de modificaciones de histonas asociadas con la transcripción activa incluyen la metilación de histonas H3K4/K20/K36/K79, acetilación de la lisina de H3 y H4, así

como H3R17me2 y H4R3me2. Estas modificaciones a menudo son reconocidas por proteínas "lectoras" especializadas que reclutan otros factores para crear una estructura de cromatina abierta. Contrario a ello las modificaciones que conducen represión transcripcional, incluyen la metilación de H3K27, la desacetilación de H3 y H4 , el H3R2me2 y H4R3me2 (Lee & Kim, 2022).

Las modificaciones de histonas, junto con la metilación del DNA pueden afectar la inactivación de genes supresores de tumor. Se ha encontrado en NSCLC y lesiones displásicas bronquiales pre-invasivas acetilación excesiva de H4K5/H4K8 y pérdida de trimetilación de H4K20 (Brzeziańska *et al.*, 2013). Así mismo, se han detectado niveles globales más bajos de H4K20me3 en lesiones precursoras y es particularmente común en cánceres escamosos; también se han asociado cambios epigenéticos en H2A (H2AK5ac) y H3 (H3K4me2, H3K9ac) en NSCLC en estadio temprano (Barlési *et al.*, 2007).

En el cáncer de pulmón las HDAC se sobreexpresan a través de distintos mecanismos, por ejemplo, SIN3A forma parte de un complejo represor de HDAC y está regulado a la baja en NSCLC. Por otro lado, a través de proteínas con dominios de unión a metilo se pueden reclutar enzimas modificadoras de histonas como MBD, KAISO, MeCP2 que impacta la modificación en el código de histonas (Brzeziańska *et al.*, 2013).

### **1.17. Regulación y Factores de transcripción**

Los factores de transcripción (TF) cumplen funciones importantes durante la embriogénesis y la diferenciación de tejidos ya que tienen un papel importante en la especificación de la identidad celular, lo que sugiere que dichas proteínas regulan conjuntos muy específicos de genes blanco. Además, los TF pueden activar o reprimir la transcripción, según el contexto ya que muchos de estos no funcionan de forma aislada, sino que interactúan con otras proteínas en complejos (Lambert *et al.*, 2018).

Estos reconocen secuencias de DNA específicas de 6 a 12 pares de bases (pb) su

participación en la regulación génica implica el reclutamiento de coactivadores transcripcionales o correpresores, modificadores de histonas y proteínas remodeladoras de nucleosomas en un dominio local en la cromatina. En consecuencia, las regiones reguladoras de genes funcionales, incluidos los promotores y enhancers, los elementos represores de genes y los sitios de arquitectura del genoma, ensamblan grandes complejos (Reiter *et al.*, 2017).

También existen factores de transcripción cuyo sitio objetivo se encuentra en las regiones de heterocromatina (resistente a nucleasas) promoviendo la accesibilidad para el reclutamiento de otras proteínas como cofactores y remodeladores de cromatina, por lo que juegan un papel principal en la programación y reprogramación celular. Para unirse al DNA nucleosomal, los TF deben competir con los nucleosomas o interactuar con ellos para acceder a sus sitios. Existen diferentes clases de factores de transcripción como los denominados pioneros que tienen distintas características de unión a nucleosomas. Por ejemplo ciertos factores básicos de cremallera de leucina (bZIP) y hélice-bucle-hélice básica (bHLH), particularmente aquellos que rodean más de 180° de la circunferencia del DNA, se unen a sitios en el borde del nucleosoma. Los factores de transcripción de dedo de zinc, homeodominio y caja de cabeza de horquilla se unen al ADN periódicamente alrededor del nucleosoma, mientras que los factores Sox (que contiene la caja HMG) y RFX5 se unen cerca del eje central del nucleosoma (Zaret, 2020).

En la transformación maligna la expresión de diversos TF que participan en procesos que especifican la identidad celular, los patrones del desarrollo embrionario y el control de diversas vías celulares se encuentra alterada, por lo que las células sufren un proceso de reprogramación que trata del restablecimiento del programa transcripcional por lo que transitan dinámicamente a estados de pluripotencia y diferenciación (Suvà *et al.*, 2013). Algunos de estos factores clave están involucrados en el desarrollo embrionario y juegan un papel crítico en la carcinogénesis pulmonar, como *OCT4* que se encuentra sobreexpresado y es un regulador en el mantenimiento de la pluripotencialidad de



las células troncales embrionarias (Jen *et al.*, 2017, p. 4) y SOX2 (Sholl *et al.*, 2010) que forma heterodímeros con OCT4, Nanog y KLF4 (Hanna *et al.*, 2010)

Algunos otros TF reguladores del desarrollo embrionario pertenecen a los genes Homeobox y se han relacionado con la progresión del tumor y la variación en la respuesta clínica. Los genes HOX son miembros de la principal familia de genes Homeobox de TF que codifican un homeodominio de unión al DNA de hélice-girohélice de 61 aminoácidos altamente conservado (Brotto *et al.*, 2020). Los genes HOX están organizados en cuatro grupos, cada uno mapeado en diferentes cromosomas: HOXA en el cromosoma 7, HOXB en el 17, HOXC en el 12 y HOXD en el 2 (Holland *et al.*, 2007).

Estos también se expresan en la edad adulta, por lo que se sugiere que continúan desempeñando un papel en la identidad celular para el mantenimiento de los tejidos y la renovación de las células troncales; diversos estudios han demostrado la desregulación de los genes HOX en distintos tipos de cáncer como el de mama, glioma, osteosarcoma y cáncer de tiroides (H.-P. Li *et al.*, 2013).

Desde hace tiempo se ha reportado amplia metilación en islas CpG de los genes HOX, asociada a genes como: *SIX*, *LHX*, *PAX*, *DLX*, en líneas celulares de adenocarcinoma y carcinoma de células escamosas de pulmón en estadio temprano. También se demostró que la falta de metilación en las secuencias ricas en CpG se correlaciona con la presencia de la marca de activación de metilación en H3K4 en el cromosoma HOXA por lo que se ha propuesto a los genes HOX como marcadores de metilación del DNA para el diagnóstico temprano (Rauch *et al.*, 2007).

Dados los antecedentes sobre la participación de los genes HOX en los procesos de transformación y probable carcinogénesis de manera histórica se han posicionado como importantes reguladores en la Biología de los tumores pulmonares. En particular, se ha demostrado como HOXC10 promueve la progresión de varios tipos de cáncer, entre ellos cáncer de mama, gliomas, osteosarcomas y tumores malignos de tiroides. De acuerdo con esto Li y colaboradores examinaron la expresión y función de HOXC10 en células del grupo

de NSCLC como NCI-H23 y A549 xenotransplantadas en ratones inmunodeficientes; demostrando que la sobreexpresión de HOXC10 se encuentra asociada con mayor proliferación, reducción de la apoptosis y aumento del crecimiento de la masa tumoral. Por el contrario, el silenciamiento de HOXC10 mediada por siRNA en células NCI-H23, es capaz de reducir la proliferación y aumentó de la apoptosis (Li *et al.*, 2020).

## **2. ANTECEDENTES**

### **2.1. Mesenchyme Homebox-2 (MEOX2)**

El factor de transcripción MEOX2, también denominado mesénquima HOX-2, mesodermal homeobox-2 o GAX (Growth Arrest-Specific) pertenece a la familia de genes HOX, localizado en el cromosoma 7 en la banda citogenética p21.2. MEOX2 conduce la diferenciación de los tejidos mesodérmicos y se expresa en distintas interfases tisulares que exhiben interacciones epiteliales-mesenquimatosas, además la expresión de MEOX2 se describió en células de músculo liso vascular y en tejido adulto, principalmente en el sistema cardiovascular, en el riñón y la placenta (Valcourt *et al.*, 2007).

MEOX 2 con homeodominio HOX de hélice vuelta hélice reconoce secuencias ricas en Adenina y Timina. Se conforma de 3 exones que dan como resultado una proteína de 304 aminoácidos con una masa de 33.6 kDa. Ya que MEOX2 pertenece al conjunto de proteínas de la familia HOX también podría ser capaz de interactuar con componentes de la maquinaria transcripcional, así como regiones que reclutan enzimas modificadoras de la cromatina (Ladam & Sagerström, 2014).

La participación de MEOX2 ha sido descrita en anomalías del desarrollo embrionario, así como en distintas enfermedades neurodegenerativas como Alzheimer donde su haplo-insuficiencia se encuentra asociada con fenotipos de enfermedad tempranos y graves; por otro lado, MEOX2 acelera la degeneración axonal en el glaucoma (Soto *et al.*, 2016). Así mismo, ha sido descrita su participación como regulador de la proliferación de células endoteliales vasculares,

así como activador transcripcional directo de los genes supresores de tumor p21 *CIP1/WAF1* y p16 *INK4a*, por lo que, se observó que el aumento de la expresión de los factores de transcripción del homeodominio MEOX2 conduce a la detención del ciclo celular y a la senescencia de células endoteliales (Douville *et al.*, 2011). Además, se ha propuesto la participación de MEOX2 en la regulación negativa de varios blancos del factor de transcripción NF- $\kappa$ B, en células endoteliales, así como en la expresión de miembros de las familias de señalización de TGF- $\beta$ / proteína morfogenética ósea (BMP) (Valcourt *et al.*, 2007).

Adicionalmente, se ha descrito la participación de MEOX2 en los procesos de carcinogénesis como en los tumores de Wilms (Ohshima *et al.*, 2009), además de la sobreexpresión de MEOX2 tanto en lesiones precursoras como en neoplasias malignas pulmonares (Ávila-Moreno *et al.*, 2014). Adicionalmente MEOX2 ha sido propuesto como un posible biomarcador en el desarrollo histopatológico pulmonar ya que su sobre-expresión se encuentra relacionada con pobre pronóstico y capacidad de resistencia al tratamiento oncológico.

Así mismo, Tachon y colaboradores demostraron el papel de MEOX2 en células troncales de glioma (GSC), responsables de la recurrencia tumoral. Mediante ensayos de silenciamiento por siRNAs se demostró el impacto de MEOX2 en la viabilidad y el fenotipo troncal de GSC. Demostrando que MEOX2 se localiza en núcleo y posee una expresión heterogénea en GSC, además observaron que la expresión de MEOX2 depende del estado de metilación de su promotor y fuertemente asociada con mutaciones en el gen IDH (Isocitrato deshidrogenasa 1). La pérdida de la función de MEOX2 se correlacionó con la diferenciación de GSC y la adquisición de características de linaje neuronal y relaciona con mayor expresión de *CDH10*, así como disminución del gen *PTK2* que codifica una proteína tirosina cinasa citoplasmática centrada en las adherencias focales; por último, de acuerdo con este estudio MEOX2 también puede aumentar la apoptosis a través de la regulación negativa de la vía PI3K/AKT (Tachon *et al.*, 2021).

## 2.2. MEOX2 en la actualidad en cáncer pulmonar

En 2014 Ávila-Moreno y colaboradores estudiaron los perfiles de aberraciones genómicas en una cohorte de 55 pacientes del grupo NSCLC, incluyendo el análisis de lesiones precursoras; donde identificaron altas frecuencias de variaciones en el número de copias (CNV) en las regiones citogenéticas 7p22.3–p21.1 y 7p15.3–p15.2; observaron la sobreexpresión de genes, como *MEOX2*, *HDAC9*, *TWIST1* y *AhR*; además encontraron que tanto *MEOX2* como *TWIST1* mostraron niveles globales de enriquecimiento significativamente bajos en las marcas de histonas represivas H3K27me3, así como aumento en el enriquecimiento de la marca de histona de activación H3K4me3 (Ávila-Moreno *et al.*, 2014); por lo que se propuso a *MEOX2* como un importante marcador tumoral de pulmón asociado con la resistencia al tratamiento de fármacos y mal pronóstico.

En este sentido Armas-López y colaboradores estudiaron mediante ensayos de inmunoprecipitación de cromatina asociada a microarreglos de DNA (ChIP-on-chip) sobre el epigenoma del cáncer de pulmón con el objetivo de perfilar a secuencias promotores de genes blanco bajo ocupación y probablemente regulados por *MEOX2*. Su análisis bioinformático permitió identificar la firma molecular de 78 genes con un FDR= 0.02 y 13 genes con un FDR= 0.01, donde destaca el gen *GLI1*, mostrando su participación en la resistencia al fármaco oncológico cisplatino, así mismo promoviendo la migración y proliferación de células de cáncer de pulmón. Así mismo, demostraron en células A549 y H1975 de cáncer de pulmón que el eje *MEOX2-GLI1* se encuentra clínicamente asociado con los índices de supervivencia global en pacientes con status de *EGFR* no-mutado versus *EGFR*-mutado (Armas-López *et al.*, 2017).

Adicionalmente, en 2022 Peralta-Arrieta y colaboradores, con base en evidencias previas de *MEOX2*, evaluaron la capacidad de resistencia a la terapia basada en *EGFR*-TKI, así como ensayos de progresión tumoral in vivo mediante el estudio de células A549 de cáncer de pulmón; demostrando que la sobreexpresión del eje *MEOX2/GLI1* promueve la resistencia tanto a cisplatino como a la terapia basado

en EGFR-TKI, mediante la activación de la vía de EGFR. Mientras que, tanto MEOX2 como GLI1 se encuentran involucrados epigenéticamente en la expresión del gen EGFR mediante la reducción de los marcadores epigenéticos represores polycomb-EZH2, así como la histona H3K27me3, mostrando un perfil de histonas aumentado de H3K27Ac/H3K4me3 sobre las secuencias promotoras y secuencias del super-enhancer del gen (Peralta-Arrieta *et al.*, 2022).

Por último Schönrock A. y colaboradores en 2022 demostraron que la fosforilación de ERK es dependiente de la sobreexpresión de MEOX2 en glioblastomas (GBM) derivados de pacientes e implantados en ratones. Así mismo, identificaron que el residuo Ser 155 es el sitio responsable de fosforilación dependiente de ERK upstream del dominio homeobox de MEOX2, la sustitución de Ser 155 tuvo efectos en los niveles de la proteína de MEOX2, así como alteración en su localización a nivel nuclear. Así mismo, observaron aumento de la proliferación celular asociada con la sobreexpresión de MEOX2 en conjunto con la pérdida de *p53* y *PTEN* en modelos organoides cerebrales de gliomas malignos humanos (Schönrock *et al.*, 2022).

### **2.3. Regulación post-transcripcional de MEOX2 en cáncer pulmonar**

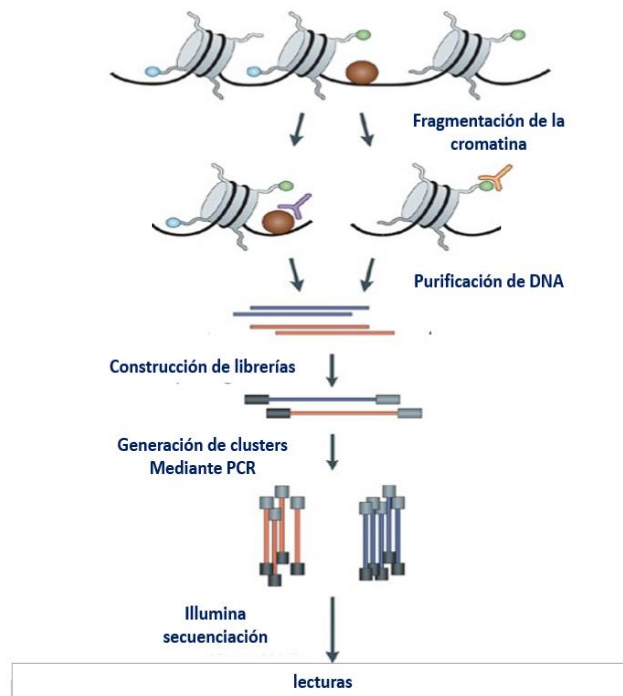
La expresión de MEOX2 se encuentra regulada post-transcripcionalmente a través de microRNAs (miRNAs) codificados a partir de genes no-codificantes capaces de unirse a secuencias idénticas o alta homología a regiones genéticas 3'-UTR del RNAm dando como resultado su inhibición de la traducción o degradación del mRNA (Chen & Gorski, 2008). A este respecto, Chen y colaboradores en 2010 demostraron como miR-221, promueve aumento en la expresión de MEOX2 a través de la regulación negativa de SIP1/ZEB2, miembros de la familia ZEB como factores de transcripción tipo dedos de zinc, caracterizados por un homeodominio separado por la estructura de dedos de zinc altamente conservados (Chen *et al.*, 2010).

Adicionalmente se ha demostrado que la inhibición de miR-301 en células de cáncer pulmonar A549 conduce a la disminución de la expresión del gen *SKA2*, a

través de un mecanismo de retroalimentación con MEOX2 afectando la vía ERK/CREB, dicha inhibición de miR-301 y SKA2 resultó en aumento del índice mitótico y disminución en la tasa de proliferación mediante ensayos de formación de colonias, lo cual puede correlacionar con el proceso de tumorigénesis pulmonar (Cao *et al.*, 2010).

#### **2.4. Análisis del epigenoma en cáncer pulmonar por ensayos de ChIP-Seq y bioinformática**

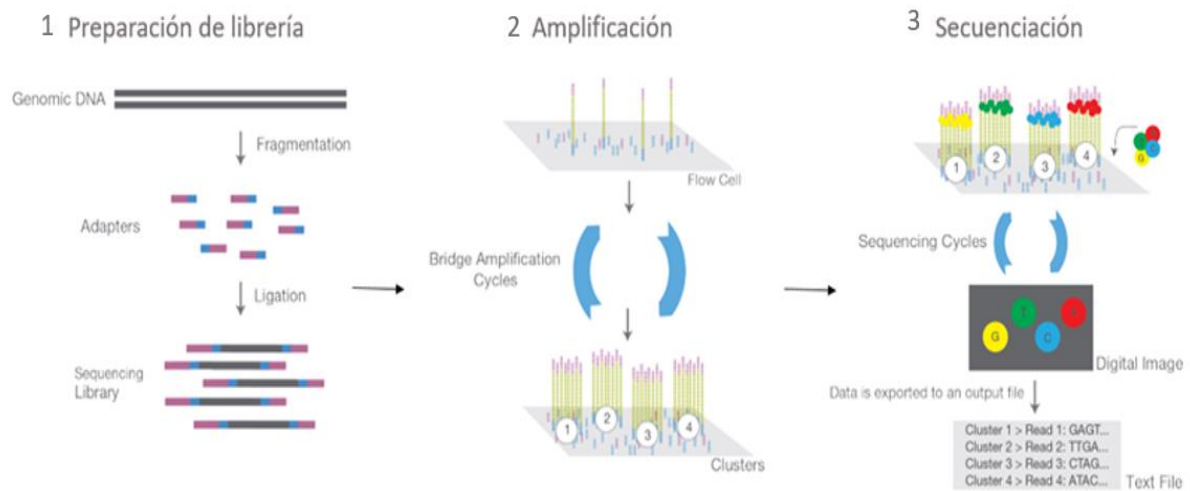
El desarrollo de ensayos de inmunoprecipitación de la cromatina (ChIP) ha resultado una técnica ampliamente aceptada y robusta para identificar regiones citogenéticas y genéticas discretas del genoma, potencialmente asociadas funcionalmente con la presencia y/o posicionamiento de proteínas específicas a lo largo del genoma. Dicha técnica comienza con el entrecruzamiento irreversible de tales proteínas a la cromatina mediante el uso de formaldehído, cuya reacción química se dirige principalmente hacia el grupo amino de la lisina de proteínas y los ácidos nucleicos. Posteriormente, las células se lisan y los fragmentos de cromatina se aíslan del material nuclear mediante fragmentación por sonicación para obtener fragmentos discretos del DNA del orden de 100 a 500 pb o mediante digestión por nucleasas; por último, se lleva a cabo el proceso de inmunoprecipitación empleando anticuerpos específicos en contra de la proteína de unión al DNA, donde se invierten los entrecruzamientos para poder analizar el DNA inmunoprecipitado mediante distintos métodos como la secuenciación de los fragmentos de cromatina obtenidos como resultado del método de ChIP-Seq (figura 2.) (Mundade *et al.*, 2014).



**Figura 3. Inmunoprecipitación de la cromatina seguida de secuenciación ChIP-Seq.** El proceso ChIP identifica la interacción DNA-proteína o los nucleosomas modificados de interés utilizando un anticuerpo específico. El DNA inmunoprecipitado se purifica y se secuencia mediante distintas plataformas existentes como Illumina, las cuales utilizan adaptadores comunes que se ligan al DNA de ChIP y se generan amplicones agrupados clonalmente (Tomada y modificada de Park, 2009).

Así mismo, existen distintos métodos de secuenciación o proceso de identificación del orden exacto de las bases del DNA detectadas por secuenciación masiva. Un tipo de plataforma de secuenciación masiva a nivel comercial ha sido la plataforma Illumina la cual emplea el método de secuenciación por síntesis (SBS) en 3 pasos (figura 2.); basándose en la DNA polimerasa la cual cataliza la incorporación de trifosfatos de desoxirribonucleótidos (dNTPs) marcados con fluorescencia durante ciclos secuenciales de amplificación en fuente isotérmica capaz de formar grupos o clusters. En cada punto de incorporación por ciclo, los nucleótidos se identifican por excitación con fluoróforos, una vez identificados se libera el terminador así como el fluoróforo, para permitir el siguiente ciclo de incorporación de nucleótidos, esto genera múltiples imágenes colorimétricas las cuales son convertidas en

lecturas de secuencias (Bentley *et al.*, 2008).



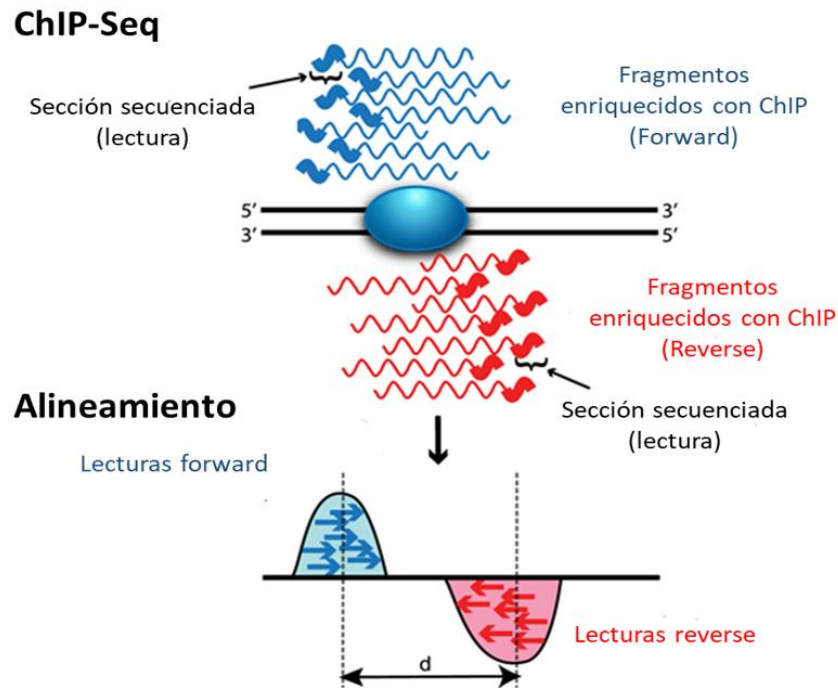
**Figura 4. Proceso de secuenciación a través de plataforma Illumina.** 1) La preparación de librerías de secuenciación a través de ligadura de adaptadores en ambos extremos de los fragmentos 5' y 3' de la muestra de DNA o cDNA. 2) Las librerías se hibridan en una celda de flujo y se amplifican en puente, dando como resultados distintos clusters para cada fragmento de la celda de flujo. 3) Durante la secuenciación se añaden nucleótidos marcados fluorescentemente a las bases complementarias del DNA en los clusters; posteriormente, el instrumento capta con láser la fluorescencia. Tomada y modificada de (Illumina, 2017).

Basado en todo lo anterior, las tecnologías de secuenciación de nueva generación (NGS por sus siglas en inglés) han permitido la lectura de millones de secuencias del DNA genómico de forma masiva y en paralelo para el estudio extenso del genoma. Una vez que se obtienen grandes cantidades de datos de lectura se requiere de su manejo y análisis a través de un proceso bioinformático y experimental para su interpretación (Levy & Boone, 2019).

Por lo tanto, los ensayos de ChIP-Seq han resultado una técnica eficaz para la identificación de la unión de factores de transcripción específicos (TFBSs), a sitios (dominios) en el DNA genómico. En este sentido, recientemente los análisis bioinformáticos de ensayos de ChIP-Seq han sido empleados en la detección de múltiples blancos genéticos implicados en procesos de malignidad oncológica (Jen



*et al.*, 2017; Namani *et al.*, 2019; Xu *et al.*, 2019), algunos de los cuales se han descrito previamente en los hallmarks del cáncer.



**Figura 5. Secuenciación masiva y análisis de la distribución de lecturas.** Los fragmentos en color azul y rojo representan los segmentos de enriquecimiento de la proteína de interés obtenidos en ChIP (óvalo azul); los fragmentos engrosados indican las regiones secuenciadas por tecnologías de lectura corta. Los fragmentos secuenciados se alinean con un genoma de referencia y se proyectan en una coordenada cromosómica (flechas rojas y azules), lo cual genera una longitud (d) dada por las crestas de los sitios enriquecidos para construir los picos representativos de la unión del TF Tomada y modificada de (Wilbanks & Facciotti, 2010).

Con base en lo anterior, la detección de regiones genómicas que se han enriquecido por la Inmunoprecipitación (concepto conocido como el llamado de picos) resulta un paso central del procesamiento de datos que derivan de ensayos de ChIP-Seq, también reconocida como la detección de señales de enriquecimiento asociado a interacciones proteicas sobre el DNA genómico que, acompañadas de pruebas de validación experimental biológica, permiten proponer nuevos mecanismos potenciales de modulación genético-transcripcional, y epigenética de la expresión en el genoma del cáncer (Nakato & Shirahige, 2017).

El estudio de muestras biológicas procesadas a través de ensayos de ChIP-Seq, arroja además de la señal real del enriquecimiento, una gran cantidad de fragmentos generados a partir de regiones genéticas no-específicas en todo el genoma, debido a los contactos aleatorios por parte de la proteína con el DNA, o bien anticuerpos con el DNA; en este sentido las lecturas secuenciadas de estos fragmentos se encuentran ampliamente distribuidas en el genoma y se consideran ruido de fondo, por lo que, son empleados secuencias del DNA control llamados controles negativos para estimar la distribución, amplitud y profundidad así como variación en el número de copias del genoma de interés, con el fin de ajustar posibles sesgos intrínsecos durante la experimentación para el estudio del genoma o bien epigenoma. Por lo que, existen distintos controles para restar el fondo como los denominados DNA-Input, que es aquella muestra de cromatina libre de Inmunoprecipitación o una muestra con inmunoprecipitación simulada mediante el empleo de anticuerpos no-específicos para detección de proteína de no-humano “uso de Abs IgGs” (Liang & Keleş, 2012). Adicional a ello, actualmente los llamadores de picos tienen modeladores internos del fondo para eliminar la señal de ruido. Por lo que se han propuesto e implementado distintos algoritmos para la optimización y mejora de la sensibilidad y precisión en la detección de picos. Algunos autores aceptan que el algoritmo óptimo puede depender del conjunto y las características de los datos a analizar mientras que otros aconsejan la combinación del resultado de diferentes métodos. Este es un punto central en el análisis ya que el enfoque utilizado puede conducir a la pérdida de verdaderos sitios de unión, y por lo tanto a la pérdida de información biológica relevante (de Boer et al., 2014)

### 3. JUSTIFICACIÓN

Los estudios de ciencia básica, preclínico y clínico del cáncer pulmonar resulta fundamental para lograr la comprensión de estrategias de atención en el manejo del paciente oncológico con miras a mejorar la calidad de vida, tomando en cuenta que los pacientes con cáncer pulmonar poseen sobrevida global a 5 menor del 18%. Dado que, este tipo de neoplasia se encuentra ampliamente relacionada con la exposición a factores de riesgo que promueven amplias aberraciones en el genoma *versus* epigenoma, capaz de afectar la progresión, malignidad y respuesta al tratamiento, por lo que, resulta importante analizar dichos cambios y/o aberraciones moleculares.

A este respecto, el factor de transcripción MEOX2 se ha propuesto como biomarcador de prognosis y respuesta a la terapia, resultando fundamental detectar y analizar sus blancos genéticos sobre el genoma empleando metodologías masivas para el estudio funcional del epigenoma, mediante la técnica del ChIP-Seq; destacando actualmente como una de las herramientas de análisis bioinformático más robustas para el análisis de gran número de datos, potencialmente libres del uso de un control de DNA de referencia para el análisis de secuenciación del genoma como se plantea con el uso del DNA-Input, empleado para el restado de fondo y obtención del llamado de picos genéticos significativos; por lo que, en el presente proyecto se propone realizar un análisis comparativo empleando datos de ChIP-Seq de nuestro grupo de investigación y el uso de datos públicos de DNA genómico de referencia como son los DNA-Input, para el estudio del posicionamiento de proteínas sobre el epigenoma como es el caso del factor MEOX2 sobre el epigenoma del cáncer pulmonar.

#### **4. PLANTEAMIENTO DEL PROBLEMA**

Se ha realizado un gran esfuerzo para determinar experimentalmente e inferir computacionalmente la unión de los TF al DNA genómico, con el objeto de asociarlos con la regulación transcripcional de distintos genes, algunos de los cuales constituyen miembros de vías de señalización celular; en este sentido, dada la evidencia de la participación de MEOX2 en la progresión, malignidad y variación en la respuesta al tratamiento oncológico, resulta importante identificar los posibles blancos génicos y sus interacciones a lo largo del genoma. Siendo importante resaltar que durante el protocolo de ChIP-Seq se sugiere el uso de controles de genomas de referencia (DNA-Input) para el restado de fondo durante la obtención de llamado de picos de fluorescencia (lecturas) producto de la secuenciación masiva del DNA genómico inmuno-precipitado. En este sentido a la fecha, existe un amplio repositorio de datos públicos de secuenciación masiva los cuales nos propusimos utilizar para el estudio bioinformático multi-comparativo. Identificando algunos blancos de MEOX2 a través de ChIP-Seq, adicionalmente desarrollar ensayos funcionales de validación experimental.

## **5. HIPÓTESIS**

El análisis bioinformático comparativo de la inmunoprecipitación de cromatina (ChIP) asociada a la secuenciación masiva del genoma asociado al factor de transcripción MEOX2, permitirá el análisis amplio del epigenoma así como la identificación de regiones de regulación epigenética involucradas en procesos oncológicos pulmonares.

## **6. OBJETIVO GENERAL**

Analizar el posicionamiento de MEOX2 sobre el epigenoma de células de cáncer pulmonar, mediante ensayos de ChIP-Seq y análisis bioinformático comparativo.

## **7. OBJETIVOS PARTICULARES**

1. Analizar la calidad de los datos de secuenciación derivados de ensayos de ChIP-Seq, y su alineamiento con base de datos públicas con genoma de referencia de cáncer pulmonar.
2. Determinar el llamado de picos de enriquecimiento asociados a MEOX2, a partir de ensayos de ChIP-Seq.
3. Realizar análisis comparativos de los datos obtenidos de ChIP-Seq con DNA-input derivados de genomas de referencia de cáncer pulmonar públicos.
4. Determinar mediante análisis bioinformático los blancos génicos de MEOX2 y la construcción de vías de señalización o procesos moleculares.
5. Realizar análisis de validación mediante ensayos de RT-qPCR sobre algunos blancos genéticos identificados para el factor transcripcional MEOX2.

## **8. MATERIALES Y MÉTODOS**

El desarrollo de ensayos de ChIP-seq para la obtención de datos de secuenciación masiva, realizados por nuestro grupo de investigación de Enfermedades Pulmonares y Epigenómica del Cáncer UNAM-INER de la Unidad de Biomedicina (UBIMED), se describe a continuación:

### **8.1. Fragmentación de la cromatina**

Se utilizó la línea celular de adenocarcinoma pulmonar A549, con silenciamiento por short hairpin (shSCR, shMEOX2, shGLI1). El cultivo celular se lavó con PBS 1X frío, posteriormente se fijaron las células con formaldehído al 1%, luego se agregó glicina a 0,125 M, y se realizaron dos lavados con PBS 1X frío, se agregó PBS y las células adherentes se rasparon y transfirieron a un tubo Falcon, el sedimento celular se lavó con PBS 1X frío, las células se resuspendieron en tampón RIPA con inhibidor de proteasas. Las muestras se sonicaron y el sobrenadante se dividió en alícuotas, posteriormente se agregó proteinasa K a la muestra, y se añadió un volumen de fenol:cloroformo, se tomó el sobrenadante y se añadieron dos volúmenes de isopropanol frío, se dejó precipitar el DNA durante la noche a -80°C. Después de la precipitación, el sedimento de DNA se lavó tres veces con etanol al 70 %, por último el sedimento se resuspendió en agua libre de nucleasas.

### **8.2. Inmunoprecipitación de la cromatina**

Después de que las muestras fueron fragmentadas, se agregó el anticuerpo correspondiente SMARCB1 (ab192864), EZH2 (ab186006) y MEOX2 (sc-81971) más perlas magnéticas, la incubación fue durante la noche a 4°C en un rotador DYNAL. Las muestras fueron lavadas con los siguientes tampones en orden: tampón bajo en sal, tampón alto en sal, tampón de cloruro de litio, tampón TE. Una vez terminados los lavados se adicionó tampón de elución ChIP, proteinasa K y tampón de unión. Las muestras se transfirieron a un tubo colector para

centrifugarlas, se desechó el sobrenadante y se agregó tampón de lavado, lo anterior se repitió 3 veces más y se agregó agua libre de nucleasas. Por último el DNA inmunoprecipitado (IP-DNA; 20 ng) se amplificó linealmente usando un GenomePlex Complete Whole Genome Amplification Kit (WGA2-50RXN) IP-DNA-WGA2; 500 ng, se cuantificó en un NanoDrop y se realizó una qPCR.

### **8.3. Construcción de bibliotecas y secuenciación masiva.**

Para la construcción de la biblioteca se evaluó el perfil de tamaño de DNA en un bioanalizador 2100. Se utilizó el kit Truseq Chip de Illumina con un total de 10 ng. Los fragmentos de DNA fueron de 300 a 500 pb; la distribución de las bibliotecas amplificadas fue evaluada mediante Tape Station 4200, cuya concentración final fue de 1,5 pM y se incluyó el 30% del control PhiX. La secuenciación se llevó a cabo en el Laboratorio de Secuenciación del “Hospital de Especialidades” con la plataforma Illumina NextSeq 500/550 High Output.

### **8.4. Análisis bioinformático**

Las lecturas utilizadas para el análisis fueron las provenientes de la línea celular A549 short hairpin Scramble (shSCR) con anticuerpos específicos para MEOX2. Para analizar dichos datos se utilizaron controles de secuenciación derivados de ensayos de CHIP-Seq. Los datos control o DNA-Input se obtuvieron de la base de datos del Centro Nacional de Información Biotecnológica (NCBI por sus siglas en inglés) considerando que fueran derivados de células A549 sin tratamiento. Los nombres asignados a cada muestra, el origen, el número de GEO que se asigna al depositar un experimento o perfil de expresión en la base de datos Gene Expression Omnibus para identificarlo, el archivo de lectura de secuencias (SRA) en el cual se guardan los datos de secuenciación sin procesar, la plataforma de secuenciación utilizada y algunas otras características se muestran en la tabla 3.

Los análisis se realizaron en una computadora Lenovo MT FOC3, con capacidad de 6MiB y sistema operativo Linux de código abierto Versión 1.0.

## 8.5. Análisis de calidad

La visión global de datos fue realizada con FastQC, una herramienta de código abierto para análisis de datos con interfaz gráfica disponible en la plataforma web Galaxy versión 22.01.rc1 (Afgan *et al.*, 2018), que puede ser consultada en la liga <https://usegalaxy.org/>. Para la limpieza de lecturas de adaptadores y de calidad se utilizaron los parámetros proporcionados por default por Trim Galore versión 0.6.7 (Krueger, 2016/2022), también incluido en Galaxy.

## 8.6. Alineamiento y análisis de calidad

Para el alineamiento y análisis de calidad se usó BWA-MEM versión 0.7.17 (H. Li, 2013; H. Li & Durbin, 2009) incluida en Galaxy contra el genoma de referencia humano hg19 GRCh37. Por último se filtraron los datos considerando una calidad de mapeo inferior a 20 ( $p > 0.02$ ) con la herramienta samtools versión 1.8 (H. Li *et al.*, 2009).

## 8.7. Obtención de picos

Los picos (regiones que se han enriquecido representando una ubicación de unión putativa para el factor de transcripción y con una puntuación asociada como indicador de significancia) de las lecturas alineadas fueron identificados con MACS versión 2.1.1.20160309.6 con los parámetros  $FDR=0.05$ ,  $mfold$  mínimo=5,  $mfold$  máximo=50 y  $band\ width=300$  (Zhang *et al.*, 2008) y findPeaks del paquete de software de código abierto HOMER v4.11 (Hypergeometric Optimization of Motif EnRichment) (Heinz *et al.*, 2010) con los siguientes parámetros: tasa de descubrimiento falso  $FDR=0.001$ ,  $p$  de Poisson=  $1.00e-04$ , pliegue sobre la entrada requerida= 4.00, pliegue máximo bajo las posiciones únicas esperadas para las etiquetas= 2.00, y normalización a 10 millones de tags asignadas por experimento.

Todo lo anterior se repitió con cada uno de los DNA-Inputs (Input1\_UW, Input2\_UTA, Input3\_NY, Input4\_NY2, Input5\_DET) obtenidos de datos públicos para realizar el restado de fondo y obtener el análisis multicomparativo. Adicional a



ello se realizó una prueba de restado de fondo algorítmica incluida en los llamadores de picos (MACS2\_MEOX2 y HOMER\_MEOX2)

### **8.8. Anotación genómica**

Para la anotación genómica y la visualización de vías de señalización, asociados al posicionamiento del factor de transcripción MEOX2, se empleó la herramienta de análisis de búsqueda de ChIP-Seq llamada ChIPseeker de Bioconductor versión 3.15 en R studio. (Yu *et al.*, 2015).

### **8.9. Análisis de vías de señalización**

Para determinar las vías de señalización asociadas al posicionamiento de MEOX2, regulando la expresión de genes blanco-detectados mediante los ensayos de ChIP-Seq, se utilizó WebGestalt (Liao *et al.*, 2019) para 2 pruebas de enriquecimiento funcional: análisis de sobrerrepresentación (ORA) con FDR de 0.05; y enriquecimiento del conjunto de genes de prueba (GSEA) con peak score y top 10, ambos con la base de datos de Kyoto (KEEG).

### **8.10. Cultivo celular**

Se utilizó la línea celular de adenocarcinoma pulmonar A549, de origen caucásico pase 9 y 10, para la validación de genes blanco-identificados (GLI1 y EGFR) en el factor transcripcional MEOX2. Las células se mantuvieron en cajas T-75 con medio RPMI-1640 suplementado con suero fetal bovino al 10% en condiciones de cultivo de 37°C, 5% de CO<sub>2</sub> y humedad relativa; a continuación, se tomó una alícuota de células y se resuspendieron en 5 mL de medio RPMI; posteriormente, se les añadió azul de tripano, se realizó el conteo en cámara de Neubauer y se obtuvo un total de un millón para la purificación de RNA y/o proteínas.

### 8.11. RT-qPCR

Para la purificación de RNA total se utilizó el reactivo TRIzol (Invitrogen, California, USA). La síntesis de cDNA se generó a partir de 5 µg de RNA total utilizando el kit RevertAid H Minus First Strand cDNA (Vilnius, Lithuania).

**Tabla 2.** Diseño de oligos para RT-qPCR.

Gen	Secuencia (5'-3')	Tamaño	TM Exp.
GAPDH	CCCCGGTTTCTATAAATTGAGC GCCCAATACGACCAAATCC	119pb	55°C
GLI1	CCAGGAATTTGACTCCCAAG GGCTTTGAAGGGCCTCAG	125 pb	55°C
EGFR	GCGTCTCTTGCCGGAATGT GGCTCACCTCCAGAAGGTT	70 pb	55°C
MEOX2	TCTGGGACCACCTTCTTTTG CCACCACCTCTGTCACTTT	219 pb	55°C

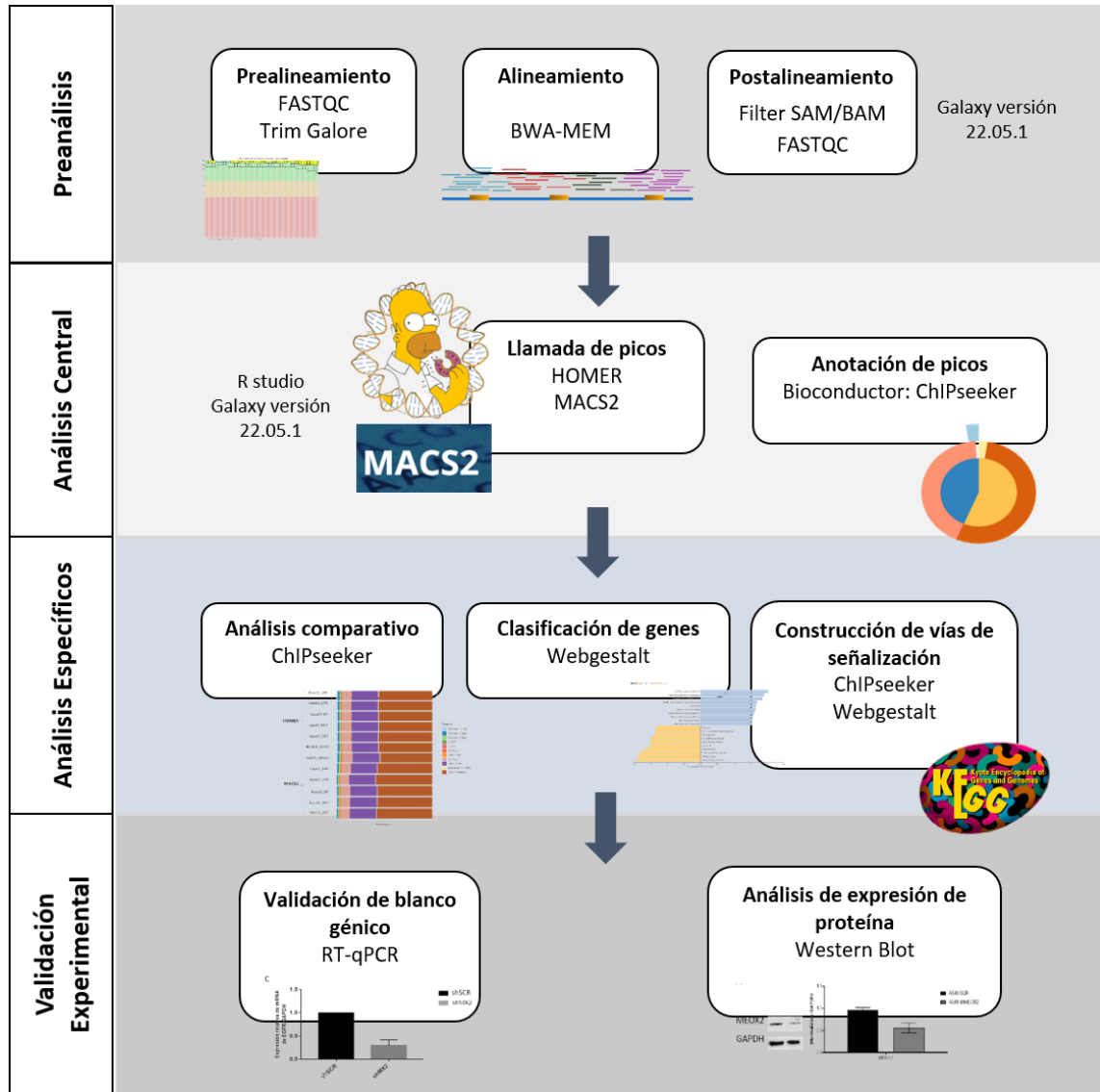
Las condiciones de amplificación fueron las siguientes: desnaturalización inicial a 95°C durante 10 min, 45 ciclos de desnaturalización a 95°C durante 10 s, alineamiento a 55°C durante 17 s y extensión a 72°C durante 30 s. Los oligonucleótidos se diseñaron usando el programa ProbeFinder versión 2.51 (Roche, Alemania) y fueron sintetizados por SIGMA-ALDRICH (EE. UU.).

### 8.12. Western blot

Las proteínas totales se extrajeron y purificaron mediante el kit comercial DC protein de Bio Rad (EE.UU.), a partir de ello se utilizaron 30 µg de proteínas totales para electroforesis vertical en geles de acrilamida al 10% a 120 volts por 120 minutos. Posteriormente se realizó la transferencia proteica a membrana de nitrocelulosa en cámara húmeda a 300 mAMP por 120 minutos, y se confirmó la transferencia con tinción con rojo de Ponceu. La membrana se bloqueó con PBS 1x Tween 20 al 0.1% con 5% de leche y se incubó con anticuerpos primarios

durante 24 hrs con los siguientes anticuerpos: GAPDH (sc-47724) 1:3000, MEOX2 (GTX55157) 1:500, EGFR (4267S) 1:3000 y GLI1 (SC20687) 1:1000. Después de la incubación se realizaron 3 lavados con PBS 1x Tween 20 al 0.1% y un cuarto lavado con agua destilada. El revelado de la membrana se realizó mediante quimioluminiscencia empleando el reactivo Immobilon Forte en cuarto oscuro, las imágenes se obtuvieron y los cambios de relación se realizaron utilizando el programa Image Studio (versión 4.0.2.1).

## 9. ESTRATEGIA EXPERIMENTAL



**Figura 6. Diagrama experimental el cual indica flujo de trabajo para el desarrollo del análisis bioinformático del posicionamiento de MEOX2 sobre el epigenoma de cáncer pulmonar, así como el desarrollo de ensayos de validación experimental.** Flujo de trabajo para el procesamiento de datos de ChIP-Seq de MEOX2 de la línea celular A549 short hairpin scramble (shSCR). Los pasos del flujo de trabajo de análisis bioinformático de datos de ChIP-seq incluyen el control de calidad, el mapeo, la detección de picos, el análisis de motivos, la anotación y la visualización. El análisis *in silico* consiste en el preanálisis, el análisis central y los análisis específicos; los resultados se validan experimentalmente mediante RT-qPCR y Western blot.

## 10. RESULTADOS

### 10.1. Preanálisis. Control de calidad de ChIP-Seq de MEOX2 y de secuencias control (input) provenientes de datos públicos.

Los protocolos de ChIP-Seq sugieren secuenciar una muestra control (DNA genómico de referencia) con el objetivo de eliminar los sesgos resultantes de la experimentación durante el análisis bioinformático; dichos controles son utilizados para estimar la distribución de fondo en cualquier posición genómica determinada, con el objetivo también de restar la variación en el número de copias o aberraciones genómicas como ocurre en el cáncer, al respecto actualmente existen una gran cantidad de datos de secuenciación públicos que pueden ser utilizados para ello, por lo que, se realizó una búsqueda en el NCBI comparando características experimentales de la muestra de ChIP-Seq de MEOX2. La información de la tabla 5 está conformada por 5 controles input en su condición experimental de secuenciación tipo *single end*, o bien *paired end* provenientes de distintos experimentos e instituciones, los cuales fueron utilizados para el presente proyecto con el objeto de evaluar las implicaciones en el análisis de predicción funcional de los datos obtenidos del ChIP-Seq de MEOX2. Dichos DNA-Input se seleccionaron al tener las características mínimas requeridas para incluirlas en el análisis de nuestra muestra experimental, es decir que provinieran de la línea celular A549 de adenocarcinoma y que no tuvieran ningún tratamiento, sin tomar en cuenta la tecnología de secuenciación. La mayoría de los algoritmos de llamada de pico requieren conjuntos de datos de control de entrada para modelar la ocurrencia de lecturas de fondo para tener en cuenta la secuenciación local y el sesgo como se mencionó anteriormente, sin embargo actualmente los avances en los modelos permiten omitir la secuenciación de un DNA-Input, ya que las herramientas cuentan con la posibilidad de un modelado de fondo algorítmico, que como se demuestra más adelante esta forma de procesamiento es comparable al uso de un DNA-Input.

**Tabla 3. Características de los datos de secuenciación de 5 controles de DNA-input provenientes del banco de datos del GEO-NCBI, y ENCODE, así como ensayos de ChIP-seq de nuestro grupo de investigación.**

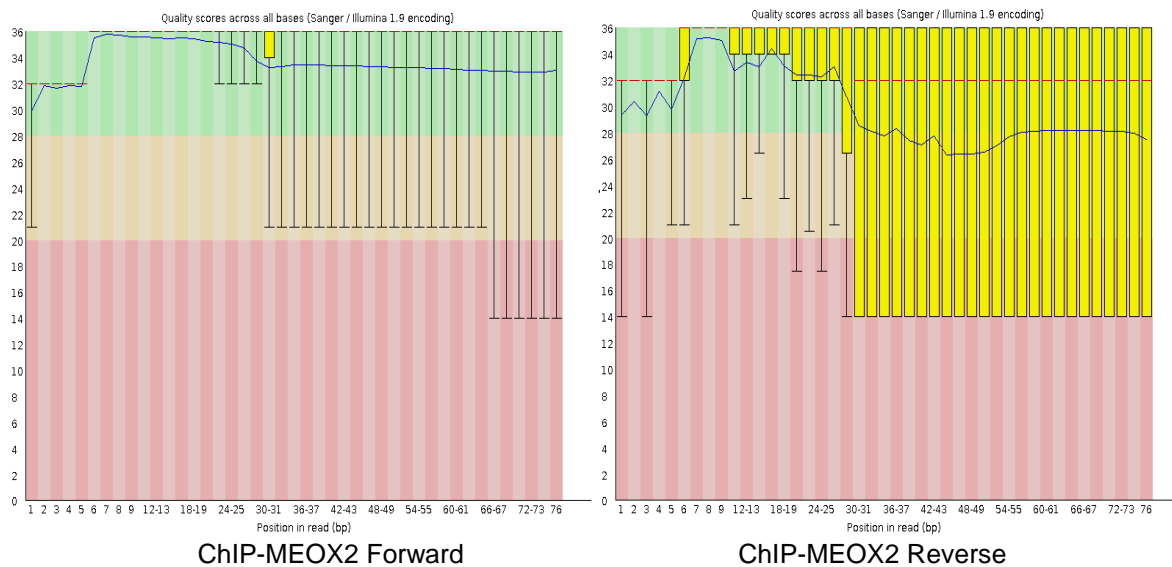
Nombre	GEO	SRA	Plataforma	Características
Input 1 UW ENCODE/ University of Washington	GSM945243	SRR577450	Illumina Genome Analyzer	Single A549 Sin tratamiento
Input 2 UT ENCODE/ UT Austin Open chromatin	GSM822307	SRR357508	Illumina Genome Analyzer	Single A549 Sin tratamiento
Input 3 R1 DNA binding of ELF1 Dittmann Laboratory, Microbiology, NYU School of Medicine	GSM4053609	SRR10043569	Illumina HiSeq 2500	Single A549 Sin tratamiento
Input 4 R2 DNA binding of ELF1 Dittmann Laboratory, Microbiology, NYU School of Medicine	GSM4053610	SRR10043570	Illumina HiSeq 2500	Single A549 Sin tratamiento
Input 5 NC Laboratory of Stem Cell Epigenomics, Oncology, Wayne State University	GSM4271455	SRR10887224	Illumina HiSeq 4000	Single A549 Sin tratamiento
MEOX2 Laboratorio 12 de Enfermedades Pulmonares y Epigenómica del Cáncer UNAM-INER. Unidad de Biomedicina (UBIMED), FES-Iztacala, UNAM			Illumina NextSeq 500	Paired A549 Anticuerpo específico para MEOX2

Nuestros resultados de secuenciación masiva sobre ensayos de ChIP-Seq fueron arrojados en archivos tipo fastq que contienen secuencias crudas libres de procesamiento, así como libres de estimación de valores de calidad. Dado que se requiere un panorama general de la calidad de los datos se realiza un reporte fastQC. En figura 7 se muestran gráficas generales que indican el control de calidad de lecturas secuenciadas obtenidas del enriquecimiento de MEOX2 en células A549 “shSCR” donde se llevó a cabo secuenciación de ambas cadenas forward y reverse.

Al respecto se muestra el control de calidad donde el *eje x* indica cada posición de las lecturas en pares de bases, así como el tamaño de los fragmentos (76 pb), mientras que en el *eje y* se muestra la escala de calidad de 0 a 36, o más que representa la probabilidad del llamado incorrecto de una base, que a su vez se divide en buena calidad (verde), calidad media (amarillo) y mala calidad (rojo); la línea azul refleja el promedio de calidad de las secuencias, mientras que la línea roja indica la mediana de la calidad en cada posición.

En figura 7, las secuencias de la cadena forward observan buena calidad a lo largo de todo el fragmento debido a que sus lecturas se encuentran por encima de

Q20, el cual es el mínimo necesario para que la llamada de base sea correcta con una probabilidad de 99.0%. Por otro lado, la calidad de las secuencias de la cadena reverse mantienen un parámetro de buena a media hasta la posición 30, a partir de la posición 31 disminuye debido a que durante el proceso de secuenciación la eficiencia de la reacción es menor conforme se agotan los reactivos (Quail *et al.*, 2008).

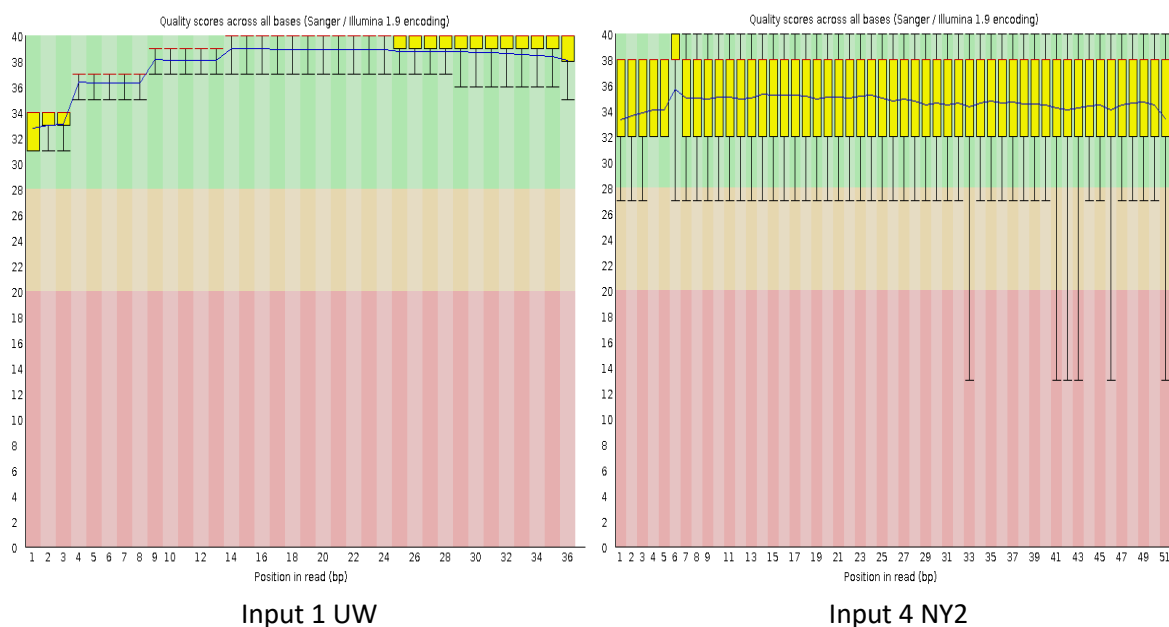


**Figura 7. Control de calidad de secuencias sentido sin procesar de MEOX2.** La calidad general de las lecturas de la cadena forward (gráfica izquierda) se encuentra entre Q30 (probabilidad de que la llamada de base sea correcta 99.9%) y Q36 (probabilidad de 99.9998 %). Las lecturas sin procesar de la cadena reverse (gráfica derecha) revelan que de la posición 1-30 están por encima de Q20 (probabilidad de 99.0%) y en adelante la calidad disminuye.

En la figura 8 se muestra la gráfica de control de calidad del DNA-input 1 UW obtenido del proyecto ENCODE de la Universidad de Washington. Se observa que todas las secuencias poseen buena calidad, por encima de Q30; se comparan con las de un proyecto independiente DNA-input 4 NY2, con menor calidad pero encima de Q20, que es el estándar aceptado con 99.0% de confiabilidad sobre que la llamada de base sea correcta.

Con estos gráficos se concluye que la calidad de las lecturas tanto de los controles DNA-Input como la secuencias derivadas de ensayos de CHIP-Seq de MEOX2

poseen buena calidad incluso antes del pre-procesamiento, por lo que, son óptimas para su uso en el análisis bioinformático.



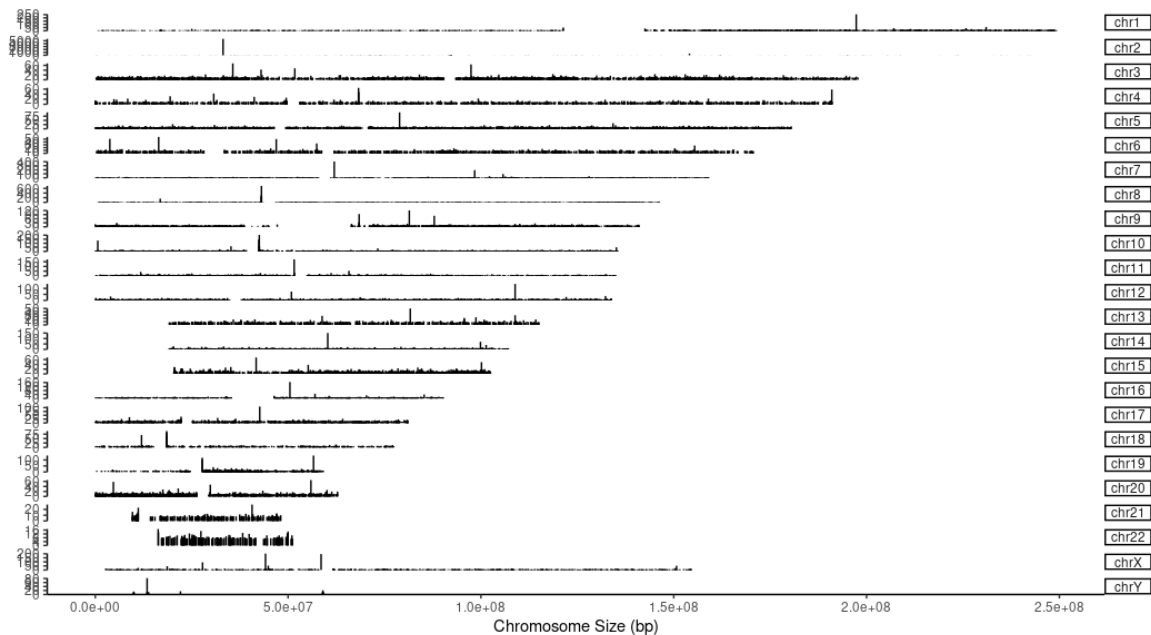
**Figura 8. Control de calidad de secuencias input 1 UW y 4 NY2.** La calidad general de las lecturas se encuentra entre Q30 (probabilidad de que la llamada de base sea correcta 99.9%) y Q36 (99.9998 %). La calidad general de las lecturas se encuentra por encima de Q20, es decir, con 99.0% de probabilidad de que la llamada de base sea correcta.

## 10.2. Análisis central: Obtención del llamado de picos (Peak calling)

Posterior al preprocesamiento el cual permite mejorar la calidad de alineamiento de secuencias amplificadas tomando como base el genoma de referencia de humano versión Hg19, se emplearon 2 modelos (Algoritmos) para el llamado de picos con el objetivo de identificar las regiones de enriquecimiento en el genoma. Los algoritmos clasifican los picos detectados con base en la señal absoluta (tamaño de lectura) o bien por significado del enriquecimiento (Peak score, valores de  $p$  y FDR), el resultado de salida bioinformático arroja archivos en formato BED, siendo a partir de ellos como se realiza el análisis sobre el enriquecimiento de los picos detectados a lo largo de todo el genoma.

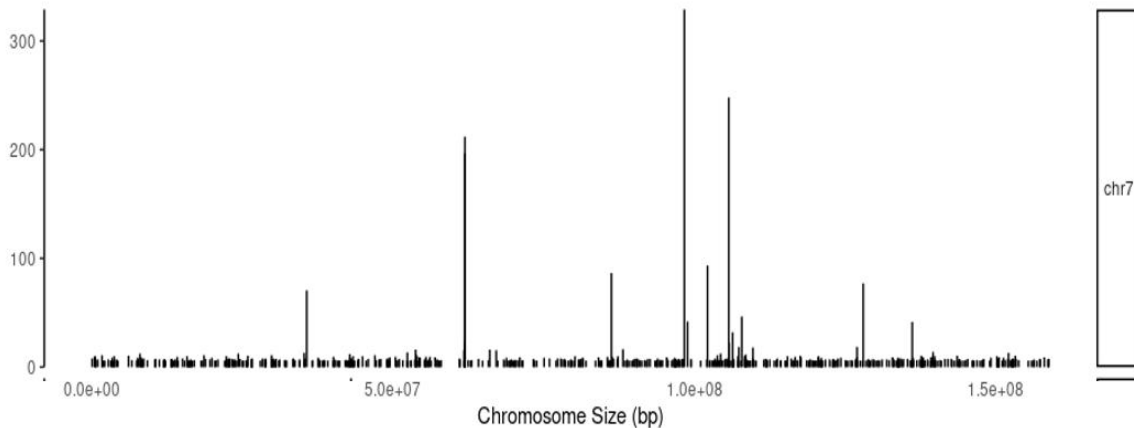


En figura 9 se muestra la cobertura a gran escala de los picos de MEOX2 a lo largo del genoma indicando su posible interacción. En el *eje x* encontramos la posición en pb de los cromosomas, así como en el *eje y* la escala de enriquecimiento, mientras que del lado derecho el nombre del cromosoma correspondiente. Se observa que los picos de MEOX2 cubren gran proporción del genoma destacando algunas regiones mayormente enriquecidas. Este resultado sugiere una amplia interacción del factor de transcripción a lo largo de todo el genoma, incluyendo los cromosomas sexuales.



**Figura 9. Interacción del factor de transcripción MEOX2 a lo largo de todo el genoma.** Se observa la ocupación del TF MEOX2 en los cromosomas humanos incluyendo los sexuales, obtenida mediante el software ChIPseeker de Bioconductor.

Las herramientas bioinformáticas permiten analizar distintas escalas del genoma. En la figura 10 se muestra con mayor resolución la heterogeneidad en la cobertura de MEOX2 sobre el cromosoma 7, siendo evidentes regiones de mayor, así como menor enriquecimiento, y por lo tanto mayor o menor ocupación de MEOX2 sobre el genoma, como parte del epigenoma. Además, algunos picos enriquecidos de MEOX2 se detectan sobre la región del locus de MEOX2.

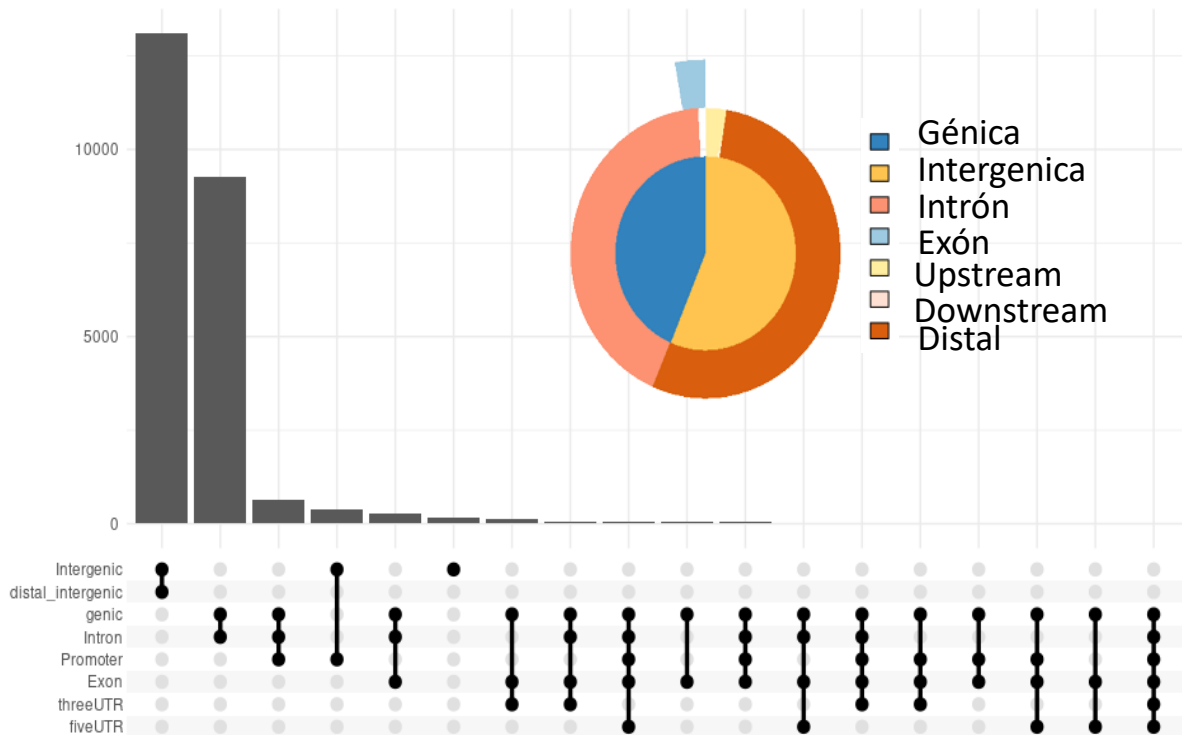


**Figura 10. Posicionamiento del factor transcripcional MEOX2 sobre el cromosoma 7.** El eje x indica la posición de pb a lo largo del cromosoma y el eje y lado izquierdo una escala de enriquecimiento.

### 10.3. Anotación genómica

La anotación genómica consiste en asociar los picos enriquecidos a los genes identificados por secuenciación, por lo tanto, durante este proceso se asignan características genómicas espaciales, estructurales y funcionales de todo el genoma. En la figura 11, se muestra la distribución de los picos de enriquecimiento de MEOX2 sobre cada uno de los elementos o segmentos lineales sobre el genoma, así como sus intersecciones. El histograma muestra en el eje “x” la matriz de elementos genéticos compartidos, mientras que en el eje “y” la abundancia del número de picos presentes en los diferentes rangos de regiones genéticas específicas. Se muestra que la ocupación de MEOX2 mayoritariamente se encuentra en el conjunto de regiones intergénicas, así como intergénicas distales seguidas de las regiones génicas e intrónicas; por otro lado, existe menor ocupación de MEOX2 en el sub-conjunto de regiones génicas-intrón-promotor, lo cual, resulta un fenómeno poco convencional en general para el posicionamiento de factores de transcripción en condiciones no tumorales.

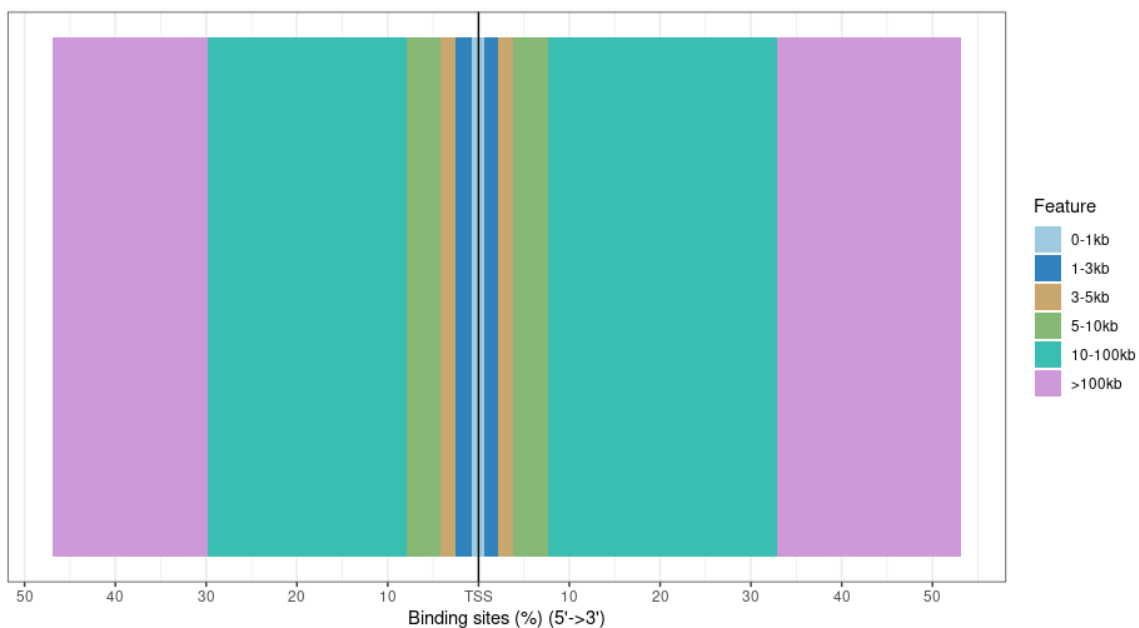
La gráfica circular intenta distinguir la anotación completa con la superposición de las regiones cromosómicas, de las cuales la ocupación de MEOX2 es dominante sobre las regiones intergénicas distales, en contraste de las regiones intergénicas con superposición de regiones génicas downstream.



**Figura 11. Posicionamiento de MEOX2 sobre el Epigenoma del cáncer pulmonar. La distribución-ocupación de MEOX2 ocurre principalmente sobre regiones génicas y distales intergénicas.** Dicha ocupación de MEOX2 sobre los distintos elementos genómicos se muestran en la matriz (intervalos genómicos) que representan las superposiciones de 7 regiones genómico-cromosómicas ordenadas por tamaño; los círculos más intensos en la matriz indican los conjuntos de la intersección.

En la figura 12 se muestra la distribución porcentual de los sitios de unión de MEOX2, dicho análisis asigna la distribución de unión de MEOX2 a genes blanco con relación al TSS. Se logra observar que la mayor proporción del posicionamiento de MEOX2 se encuentra principalmente en distancias genómicas

distales entre 10 a 100 kb (50%), seguido de regiones mayores a 100 kb (40%), mientras que, menos del 10% en regiones menores a 10 kb, tanto en dirección downstream, como Upstream. Dicho resultado representa la primera evidencia experimental y bio-informática reportada para el factor de transcripción MEOX2, indicando una gran interacción con los elementos distales intergénicas en el genoma, donde podrían alojarse o representar como hipótesis secuencias enhancers o bien secuencias super-enhancers.

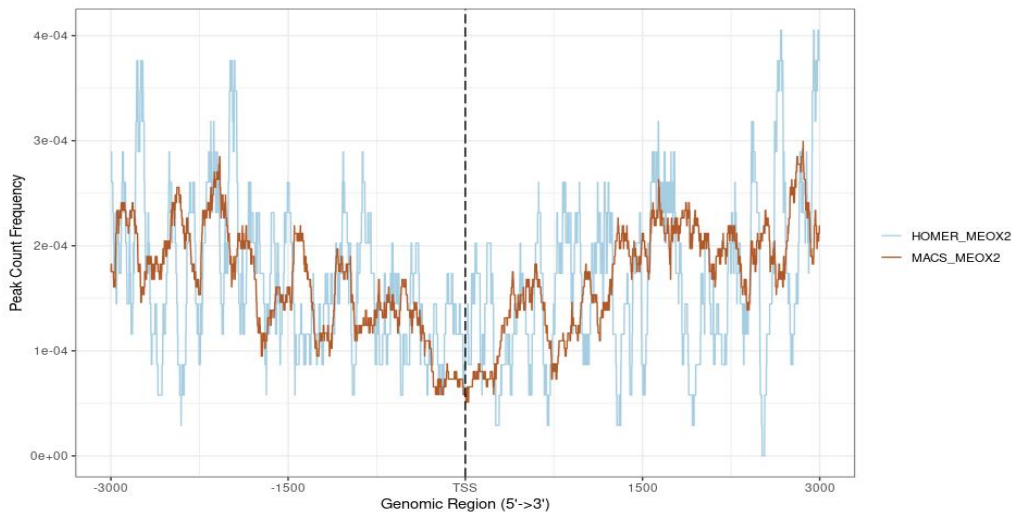


**Figura 12. Predicción bioinformática de unión de MEOX2 sobre regiones genómicas principalmente distales al TSS.** La distribución de los sitios de unión de MEOX2 con relación al sitio de inicio de la transcripción es mucho mayor conforme se aleja del TSS. En el eje x se muestra el porcentaje de la distribución de las posibles uniones del MEOX2 con relación al TSS de blancos genéticos. El código de color representa el rango en kb como se indica a la derecha.

#### 10.4. Análisis bioinformático multi-comparativo

El análisis comparativo (Ver figura 13) de los picos de fluorescencia debidos al posicionamiento de MEOX2 procesados bio-informáticamente libres del uso de DNA-Input control y basado en el uso comparativo de algoritmos computacionales conocidos como Model-based Analysis of ChIP-seq (MACS) e Hypergeometric

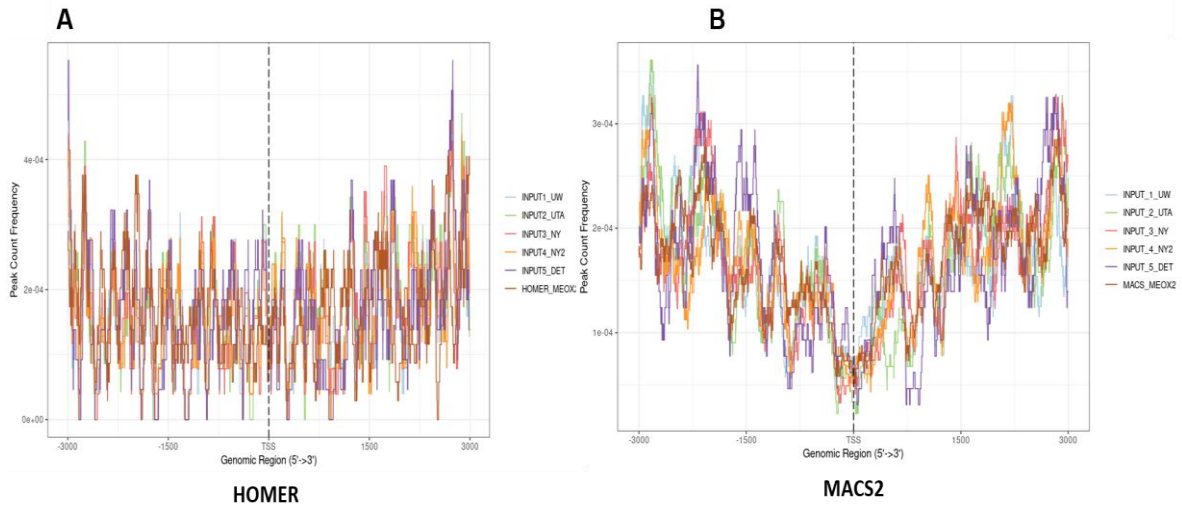
Optimization of Motif EnRichment (HOMER), con la siguientes etiquetas: MACS\_MEOX2, y HOMER\_MEOX2, muestra un patrón de enriquecimiento diferente entre cada modelo bioinformático a lo largo de las regiones genómicas indicadas en el eje “x” tomando como base el sitio de inicio de la transcripción señalado en línea punteada (TSS, por sus siglas en inglés). Dicha región fue de nuestro interés debido a que nuestros antecedentes en la ocupación de MEOX2 anlizan la región promotora de los genes blanco. Podemos observar menor dispersión de los picos generados a partir del análisis comparativo empleando el algoritmo MACS2, en contraste de los picos obtenidos empleando el algoritmo HOMER. Adicional a ello, en ambos análisis por algoritmos la distribución de sitios de unión estimados de MEOX2, revela que los picos más enriquecidos se posicionan en las regiones más alejadas al TSS, en cerca de 3,000 pb tanto en dirección downstream, como upstream. Dicho fenómeno se observa con mejor claridad, libre de control (DNA-Input), empleando MACS2, reafirmando mayor interacción (Ocupación) de MEOX2 en segmentos lejanos al sitio de inicio de la transcripción.



**Figura 13. Análisis diferencial del patrón de distribución promedio del llamado de picos de MEOX2 relativos al TSS, mediante el uso del algoritmo bioinformático MACS2 versus HOMER.** El llamado de picos libre control DNA-input se realizó mediante el uso de FindPeaks HOMER y MACS2. En el eje x se observa el perfil de picos con relación a la región genómica de 3000 pb tanto downstream, como upstream del TSS, esta región es la predeterminada por el

software. En el *eje* y tenemos la frecuencia de picos en escala logarítmica y del lado derecho el código de color para cada condición bioinformática desarrollada.

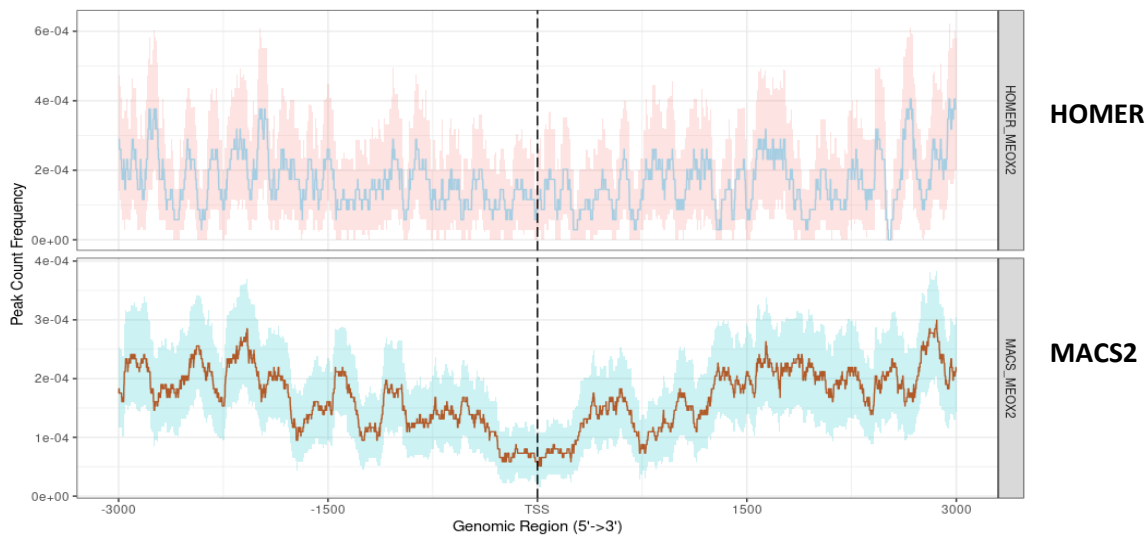
Los picos totales obtenidos oscilaron entre 8456 a 9928 para HOMER y 17763 a 28907 para MACS2. Mientras que nuestro análisis comparativo mediante el uso de 5 diferentes DNA-inputs de acceso público, nos permitió identificar un patrón de enriquecimiento comparable y homogéneo entre ellas para cada diseño algorítmico empleando HOMER, o bien MACS2 (Figura 14). Al comparar la distribución y homogeneidad del llamado de picos entre MACS2 vs HOMER se observa una diferencia evidente en el patrón de dispersión del enriquecimiento a lo largo de la región genómica analizada (Figura 14), habiendo mayor dispersión en la frecuencia de enriquecimiento determinada con HOMER a pesar de tener un valor de corte más astringente. De igual forma se identifica una tendencia al enriquecimiento conforme la ocupación de MEOX2 se aleja del TSS, confirmando lo observado con anterioridad en la distribución de los sitios de unión de MEOX2. Por lo que, podemos afirmar que el uso de distintos controles como DNA-input no afecta el patrón de enriquecimiento de los picos de MEOX2, siendo el algoritmo empleado la mayor variable bioinformática para el llamado de picos y por tanto el posicionamiento significativo de MEOX2 en el epigenoma del cáncer pulmonar. Además, observamos que ambos programas de llamada pico de uso común funcionaron bien cuando se comparó el uso de un conjunto de fondo uniforme simulado algorítmicamente (HOMER\_MEOX2 y MACS\_MEOX2).



**Figura 14. El análisis bioinformático comparativo del perfil promedio de los picos de MEOX2 sobre la región del TSS tratados con distintos DNA-inputs no muestra diferencias en el patrón de distribución relativa al TSS pero si al comparar el uso de los algoritmos HOMER versus MACS2. A)** Se observa el perfil de picos realizado mediante HOMER. En el *eje x* se encuentra la escala del perfil de picos en relación con la región genómica 3000 pb, tanto downstream como upstream al TSS, en el *eje y* se indica la frecuencia de picos en escala logarítmica, mientras que del lado derecho se indica el código de color para cada condición empleada para cada DNA-Input de acceso público. **B)** Se observa el perfil de picos realizado mediante MACS2. En el *eje x* se encuentra la escala del perfil de picos en relación con la región genómica 3000 pb downstream y upstream del TSS, en el *eje y* tenemos la frecuencia de picos en escala logarítmica y del lado derecho el código de color para cada condición.

En figura 15 se muestra el análisis comparativo del promedio de picos de MEOX2 en una región de 3000 pb tanto downstream como upstream con relación al TSS, para cada modelo algorítmico computacional utilizado HOMER versus MACS2, debido a que primero se quiso analizar la ocupación a nivel de regiones promotoras. Se observan diferencias tanto en escala, como enriquecimiento de los picos obtenidos entre HOMER versus MACS2; el primero (HOMER) muestra mayor rango de dispersión de enriquecimiento, mientras que el segundo análisis (MACS2) muestra el enriquecimiento de mayor homogeneidad sobre la región cercana al TSS. Lo anterior, demuestra que el uso de distintos llamadores de

picos modifica la sensibilidad y robustez al identificar el enriquecimiento de los picos de ocupación de MEOX2 sobre el epigenoma del cáncer pulmonar.

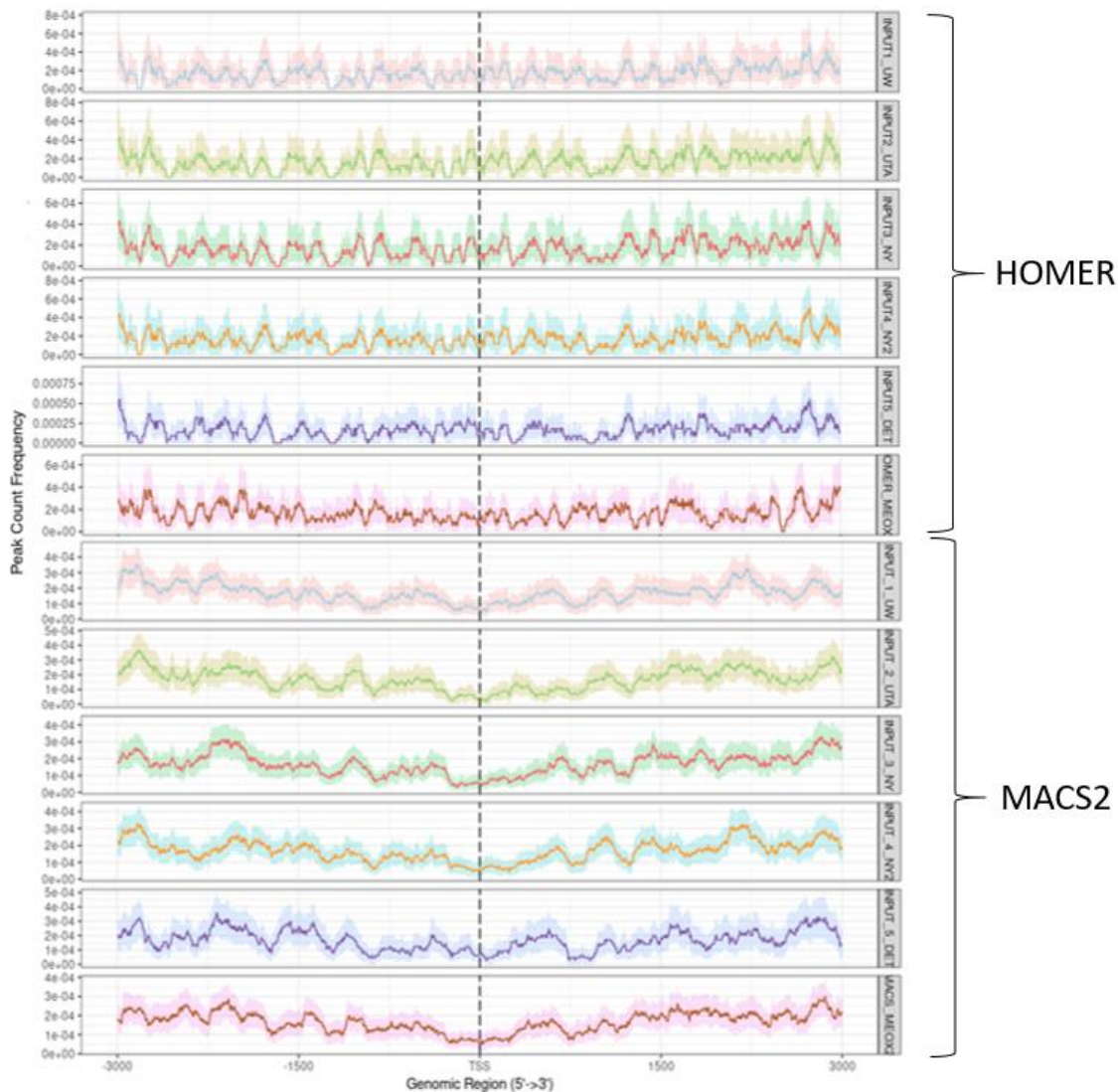


**Figura 15. Análisis comparativo del perfil promedio de los picos de MEOX2 en la región TSS.** El sombreado representa el error estándar del perfil promedio de los picos.

En figura 16 se muestra el mismo fenómeno, al comparar los picos obtenidos mediante el restado de fondo; a pesar de que su escala es distinta, el patrón de enriquecimiento es similar en las 5 condiciones de los DNA-inputs, en contraste del análisis bioinformático libre de DNA-inputs control (HOMER\_MEOX2 versus MACS\_MEOX2), tanto para HOMER como para MACS2. Sin embargo, cuando se compara el análisis obtenido a través de MACS versus HOMER, se observan diferencias en el patrón de enriquecimiento a lo largo de la región de 3000 pb con relación al sitio de inicio de la transcripción tanto downstream como upstream. Cuando se analizan los picos de enriquecimiento con relación a la escala de enriquecimiento se detecta que los picos determinados por HOMER alcanzan mayores rangos que denotan mayor dispersión, en comparación con los obtenidos a través de MACS2 los cuales oscilan en una escala homogénea entre 0e a 4e<sup>-04</sup>. Dicho patrón bioinformático se mantiene a lo largo de los análisis comparativos empleados tanto con el uso de diferentes controles de DNA-input tratados con el



mismo llamador de picos. Sin embargo, las diferencias entre los picos obtenidos a través de HOMER vs MACS2, logran arrojar distintos genes anotados, así como valor de score o puntuaciones diferentes para cada uno de los genes anotados detectados.

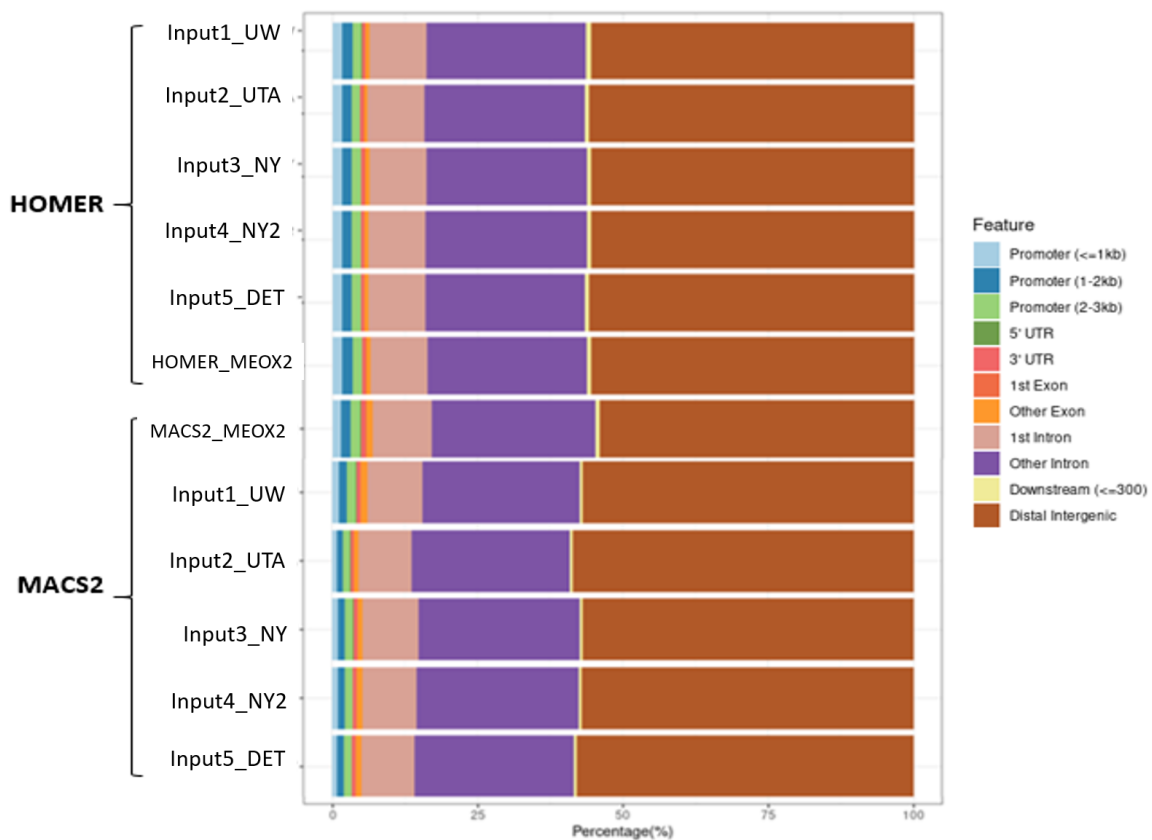


**Figura 16. Análisis comparativo del perfil promedio de los picos de ocupación de MEOX2 en la región TSS empleando distintos controles de DNA-input, y algoritmos empleados para el llamado de picos.** Los primeros 6 *tracks* son derivados del procesamiento de llamado de picos mediante HOMER, mientras que los siguientes 6 a través de MACS2. El sombreado representa el error estándar del perfil promedio de los picos.

Por otro lado, la anotación de picos observada en la Figura 17 muestra que el uso de controles de DNA-input para el restado de fondo y el procesamiento con

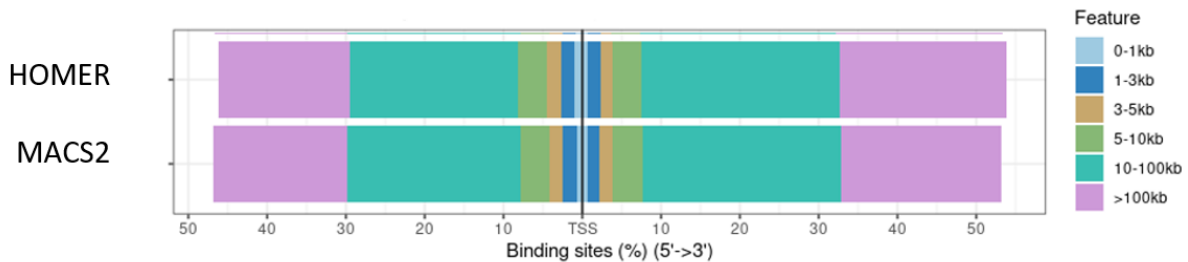
distintos modelos computacionales para el llamado de picos no modifica el análisis a pesar de haber obtenido un número total de picos distinto para cada condición DNA-input, pudiendo observar que el patrón de posicionamiento de MEOX2 se mantiene en todas las condiciones, concluyendo que se posiciona principalmente en las regiones distales intergénicas (café), seguido por secuencias de intrones (morado), secuencias de primer intrón (rosa), mientras que en mucho menor proporción en regiones promotoras de 0 a 2 kb (azul).

Cabe mencionar, que es posible detectar diferencias en el porcentaje de los picos anotados en regiones intrónicas a través del análisis realizado por MACS2 para los DNA-inputs 1 UW, y 2 UTA, así como mediante el modelo computacional de MACS2. Estas diferencias aunque pequeñas pueden resultar significativas al momento de asignar la identificación de los genes anotados y clasificar la función o vía de señalización involucrada como se demuestra más adelante.



**Figura 17. El análisis con un control input y distintos modelos computacionales para el restado de fondo y la llamada de picos no modifica la distribución de los enriquecidos de MEOX2.** Los primeros 6 tracks son derivados del procesamiento de llamado de picos mediante HOMER mientras que los siguientes 6 con MACS2. En el eje x se muestra el porcentaje de la distribución del TF en las regiones genómicas. El código de color asignado a cada región genómica, así como el rango de kilobases para cada una de ellas se muestra a la derecha.

Adicional a ello, se llevó a cabo análisis comparativo entre la distribución de los sitios de unión de MEOX2 con relación al TSS con el objetivo de identificar diferencias en la obtención de picos anotados entre el llamado a través de HOMER o MACS2; al respecto se confirmó que el posicionamiento de MEOX2 se encuentra principalmente en regiones mayores a 100 kb, seguido de regiones genéticas en el rango de 10 y 100 kb, representando más del 50% de todas las regiones identificadas, mientras que menos del 10% en regiones que van en un rango de 0 a 10 kb señaladas en color verde, amarillo y azul, como se indica en la figura 18, para ambas condiciones HOMER versus MACS2.



**Figura 18. Análisis bioinformático comparativo que muestra la distribución de sitios de unión de MEOX2 con relación al sitio de inicio de la transcripción (TSS).** En eje x se muestra el porcentaje de la distribución de los sitios de unión del TF MEOX2 con relación al TSS. Los picos utilizados en el análisis bioinformático fueron determinados a partir del restado de fondo por default que ofrecen ambos procedimientos HOMER y MACS2, es decir no fueron empleados controles negativos de DNA-input. El código de color del rango de kilobases se indica a la derecha de la figura.

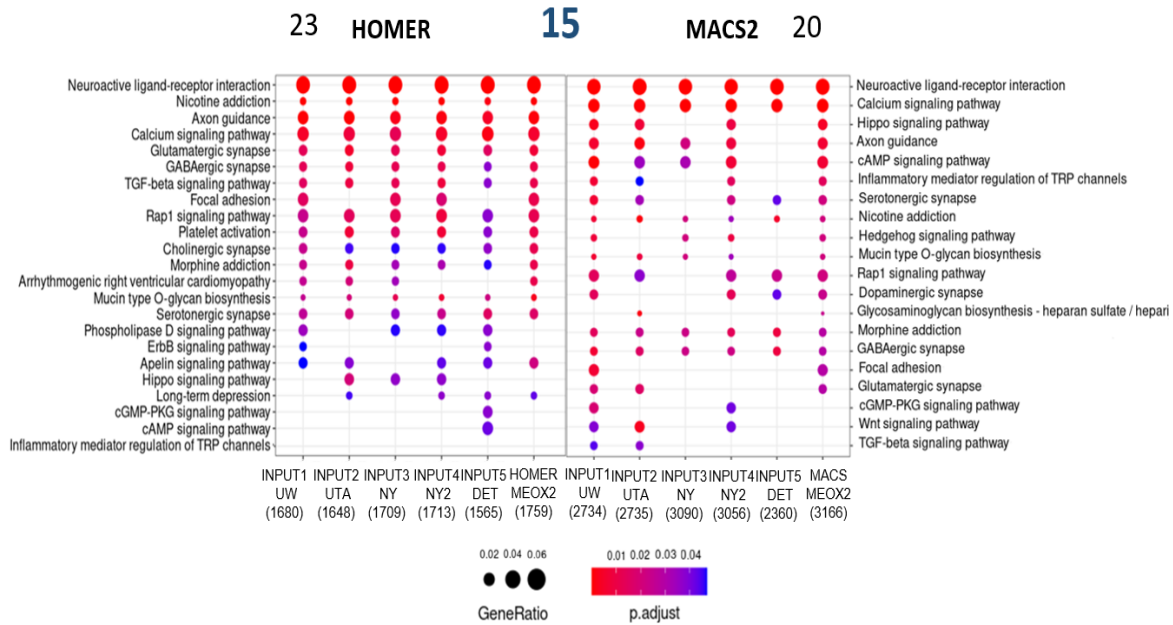
En figura 19, el análisis comparativo de vías de señalización para las 6 condiciones reveló la detección y/o enriquecimiento de vías homogéneo al

compararse dentro de cada modelo computacional HOMER o bien MACS2. Al respecto al emplear el DNA-Input UW o bien DNA-Input UTA ambos del proyecto ENCODE, se logró identificar menor cantidad de genes asociados 1680 y 1648 respectivamente, en contraste de los genes detectados al emplear el restado de fondo a través de los DNA-Inputs NY y NY2, con 1709 y 1713 genes, respectivamente. Adicionalmente, la condición con menor número de genes fue detectado al emplear el DNA-Input 4 DET con 1565 genes, así mismo destaco menor calidad de lecturas en el reporte de fastQC. En contraste, fue posible observar mejor enriquecimiento en el conjunto de genes bajo la condición del restado de fondo empleado por default por la plataforma HOMER, es decir libre del DNA-Input señalado como análisis (HOMER\_MEOX2), detectando 1759 genes.

En contraste, al comparar los genes enriquecidos mediante el algoritmo MACS2 se observó una diferencia notable en las vías de señalización enriquecidas, debido a que se encuentran en el rango de mayor número de genes detectados: 2700 y 3200 genes. Cabe destacar que la primera vía de señalización enriquecida posee el mismo valor de significancia; así como, la misma proporción de genes expresados ( $p=0.01$ , y gene ratio= 0.06), de manera independiente al DNA-Input empleado, así como al algoritmo computacional utilizado (Ver figura 19).

Al comparar las vías resultantes para cada llamador de picos se obtuvo un total de 23 vías para HOMER y 20 para MACS2 de las cuales 15 son similares para ambos llamadores de picos cuya única diferencia es su significancia de enriquecimiento.

Por otro lado, al comparar el enriquecimiento de todas las vías, entre las condiciones y de acuerdo al patrón de color y enriquecimiento (tamaño) de los puntos, las vías obtenidas de los datos procesados con los DNA-Inputs UW, UTA, NY y NY2 poseen menor significancia de enriquecimiento en contraste de los detectados al emplear el restado de fondo algorítmico (HOMER\_MEOX2 y MACS2\_MEOX2) los cuales obtuvieron un valor de  $p$  0.01 y 0.02 para la mayoría de las vías. Por otro lado, los resultados de la condición input DET poseen los valores de significancia más bajos al señalar  $p$  de 0.03 y 0.04 (Ver Figura 19).

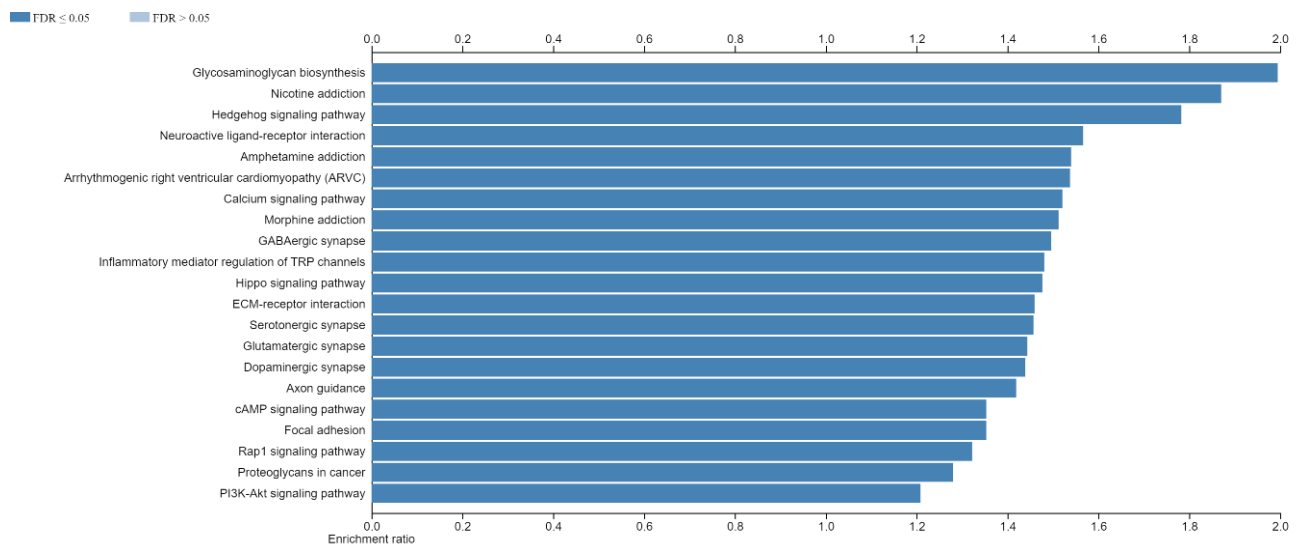


**Figura 19. Análisis bioinformático del enriquecimiento de vías funcionales de señalización bajo interacción por MEOX2.** Los puntos representan el enriquecimiento de acuerdo con un código de colores y tamaño; el rojo indica un enriquecimiento alto o significativo, mientras que el color azul de menor significancia; el tamaño de los círculos representa la proporción de genes detectados diferencialmente por cada vía de señalización.

## 10.5. Análisis de vías de señalización

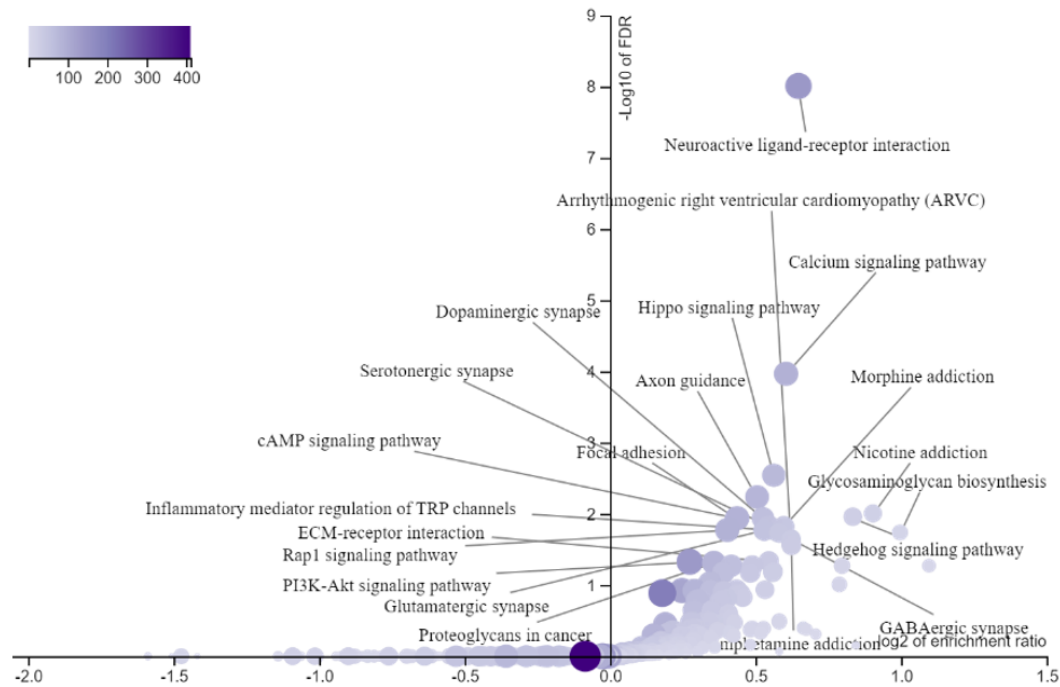
En la actualidad, los procesos de análisis bioinformático permiten el estudio del conjunto de genes asociados a un determinado proceso biológico, o bien función molecular. En este sentido las herramientas bioinformáticas para el análisis de enriquecimiento estiman la representación de categorías de genes de manera particular, o bien pertenecientes a vías de señalización a partir del empleo de distintos parámetros estadísticos. Con base en ello, se realizaron dos análisis independientes, uno de sobrerrepresentación (ORA) y otro de enriquecimiento de conjunto de genes (GSEA), todo ello, a través del empleo del programa en línea Webgestalt (Liao *et al.*, 2019) utilizando la enciclopedia de genes y genomas de la Universidad de Kioto (KEEG por sus siglas en inglés).

El análisis ORA toma en cuenta la representación de los genes ordenados en una lista de entrada para cada función biológica o vía de señalización potencialmente detectada. Los resultados obtenidos mediante el uso del DNA-Input 1 UW a través del análisis por algoritmo de MACS2 mostraron un total de 8230 genes, de los cuales 2496 representaron la integración de diferentes vías de señalización con un  $FDR \leq 0.05$  el cual se indica por el color azul fuerte (Ver Figura 20).



**Figura 20. Enriquecimiento funcional de MEOX2 mediante análisis de sobrerepresentación (ORA) mediante la plataforma de KEGG.** En el eje “x” se muestra la tasa de enriquecimiento (relación entre el número de genes observados dividido por el valor esperado). El análisis se realizó con valor de FDR de 0.05.

Adicionalmente, a continuación se muestra el análisis de ORA a través de un volcano plot (Figura 21) con el objetivo de visualizar la significancia de enriquecimiento de las vías de señalización, denotando que aquellas con tasa de enriquecimiento alto y valor de FDR bajo son más significativas. Al respecto, una de las vías de señalización encontradas fue Sonic Hedgehog, con un valor de FDR de 0.010960, así como superposición de 28 genes con respecto al conjunto completo de genes que integran la vía.



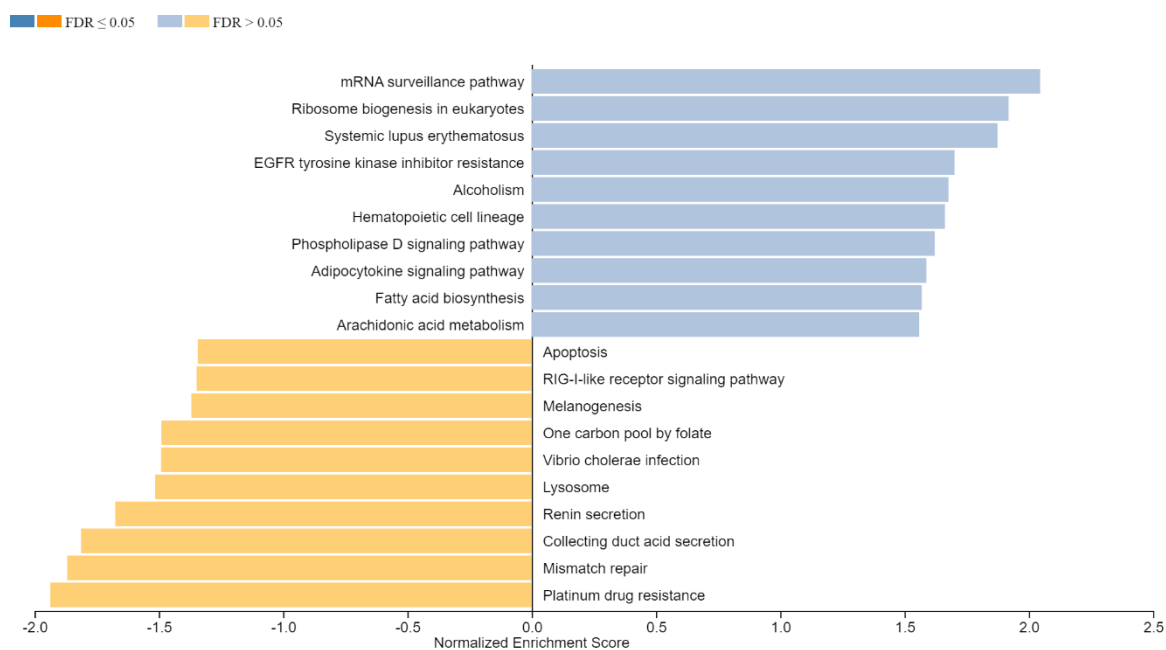
**Figura 21. Volcano plot que representa el enriquecimiento funcional de MEOX2 mediante análisis de sobrerrepresentación (ORA) por KEEG.** El eje “x” muestra la tasa de enriquecimiento y el eje “y”, así como el FDR en escala logarítmica. El tamaño y color de los círculos son proporcionales al tamaño de la categoría. Este análisis se realizó a partir del FDR con valor de <0.05.

Es importante mencionar que a pesar de que el análisis por ORA representa un análisis estándar basado en una prueba hiper-geométrica, este presenta limitantes, ya que analiza de manera aislada a cada gen, sin tomar en cuenta la estructura jerárquica correspondiente al valor significativo de su propio enriquecimiento. Con base en ello, se decidió complementar nuestro análisis mediante GSEA que analiza el enriquecimiento de genes, permitiendo evaluar los datos a nivel de conjuntos de genes y con ello determinar si estos genes aparecen en la parte superior o inferior de la lista de genes de una función o proceso biológico conocido.

El gráfico de barras de enriquecimiento muestra el score de enriquecimiento normalizado (NES) para las vías de señalización resultantes, sin embargo, todas las que aparecen en la gráfica 22, tienen un valor de FDR > 0.05 en la lista clasificada, por lo que su significancia es menor.

Las vías que aparecen con el código de color azul son aquellas donde el conjunto de genes se encuentra ampliamente enriquecido en la parte superior del listado por el contrario aquellas que aparecen en naranja poseen un menor enriquecimiento del conjunto de genes.

Una de las vías de señalización detectadas a partir del análisis GSEA fue la de resistencia a inhibidores de tirosina cinasas de EGFR, con un valor de FDR > 0.05 y valor de  $p$  de 0.026756 (Figura 22). Cabe mencionar que dicha vía de señalización había sido reportada anteriormente por nuestro grupo de investigación en cáncer de pulmón (Peralta-Arrieta *et al.*, 2022).

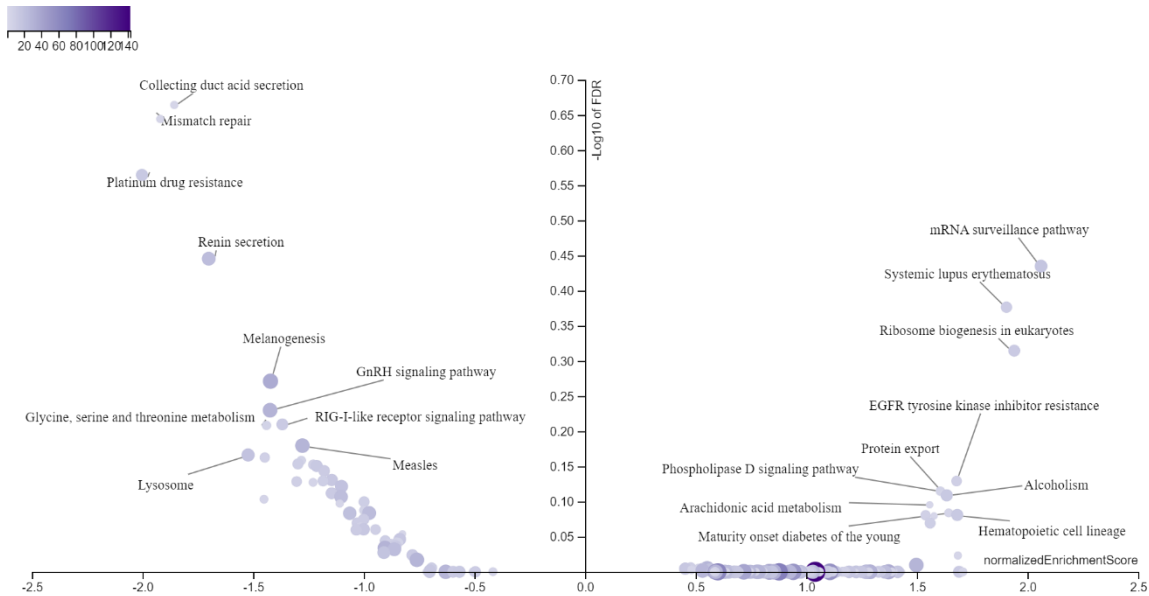


**Figura 22. Enriquecimiento funcional de procesos moleculares probablemente bajo el control de MEOX2 mediante análisis de enriquecimiento del conjunto de genes (GSEA) a través de la plataforma KEEG.** El eje “x” muestra la tasa de enriquecimiento o NES (relación entre el número de genes observados dividido por el valor esperado y el eje “y” las vías de señalización enriquecidas). El código de color muestra la significancia del enriquecimiento. El análisis se realizó a partir de un top 10.

Para este análisis bioinformático se muestra por volcano plot (Figura 23) con el objetivo de visualizar la significancia del enriquecimiento de las vías de señalización, donde aquellas con tasa de enriquecimiento normalizado alto, y valor



de FDR bajo muestran mayor significancia estadística. La vía de tirosina cinasa de EGFR obtuvo un valor de NES de 1.6796, mientras que FDR de 0.76948.

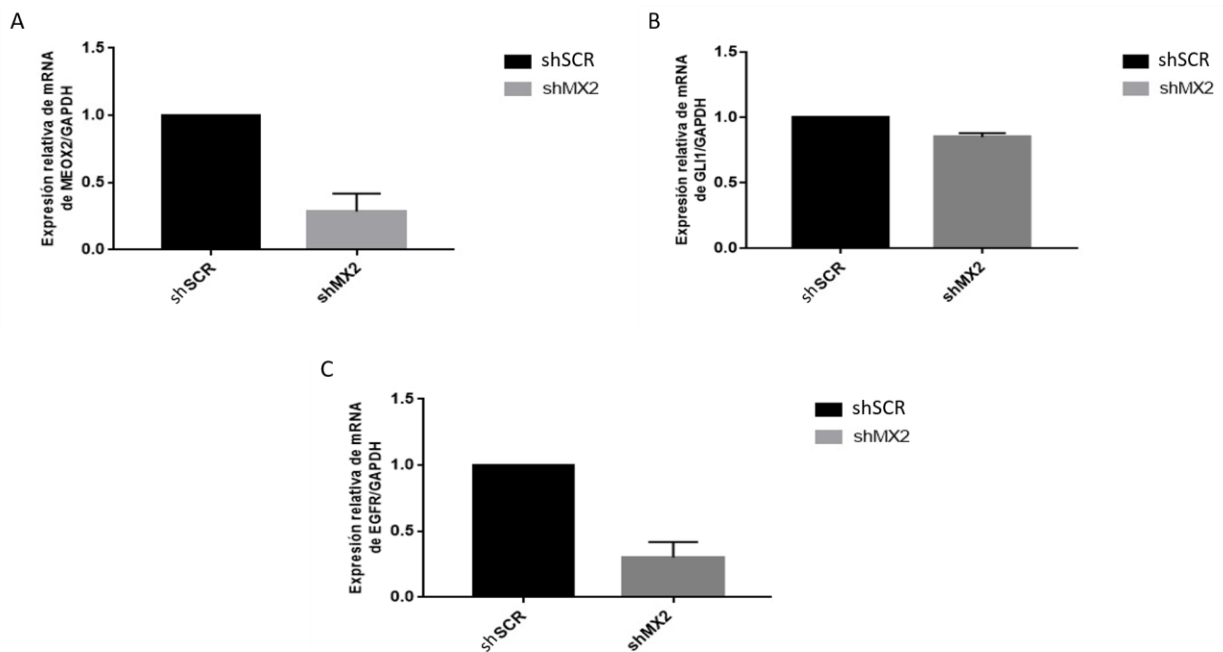


**Figura 23. Análisis de Volcano plot del enriquecimiento de vías de señalización funcional bajo el probable control de MEOX2 mediante análisis del enriquecimiento del conjunto de genes (GSEA) mediante plataforma KEEG.** El eje “x” muestra la tasa de enriquecimiento y el eje “y” el FDR en escala logarítmica. El tamaño y color de círculos son proporcionales al tamaño de la categoría. El análisis se realizó a partir del top 10.

## 10.6. Validación experimental

Dado que los análisis *in silico* de enriquecimiento mostrados anteriormente arrojaron las vías de señalización de Sonic Hedgehog (Sh) y EGFR, las cuales previamente han sido estudiadas por nuestro grupo de investigación en el contexto del cáncer pulmonar, se decidió evaluar con el objetivo de validación a 2 blancos génicos como GLI-1 miembro de la vía Sh, así como el gen EGFR siendo representativos de cada vía de señalización.

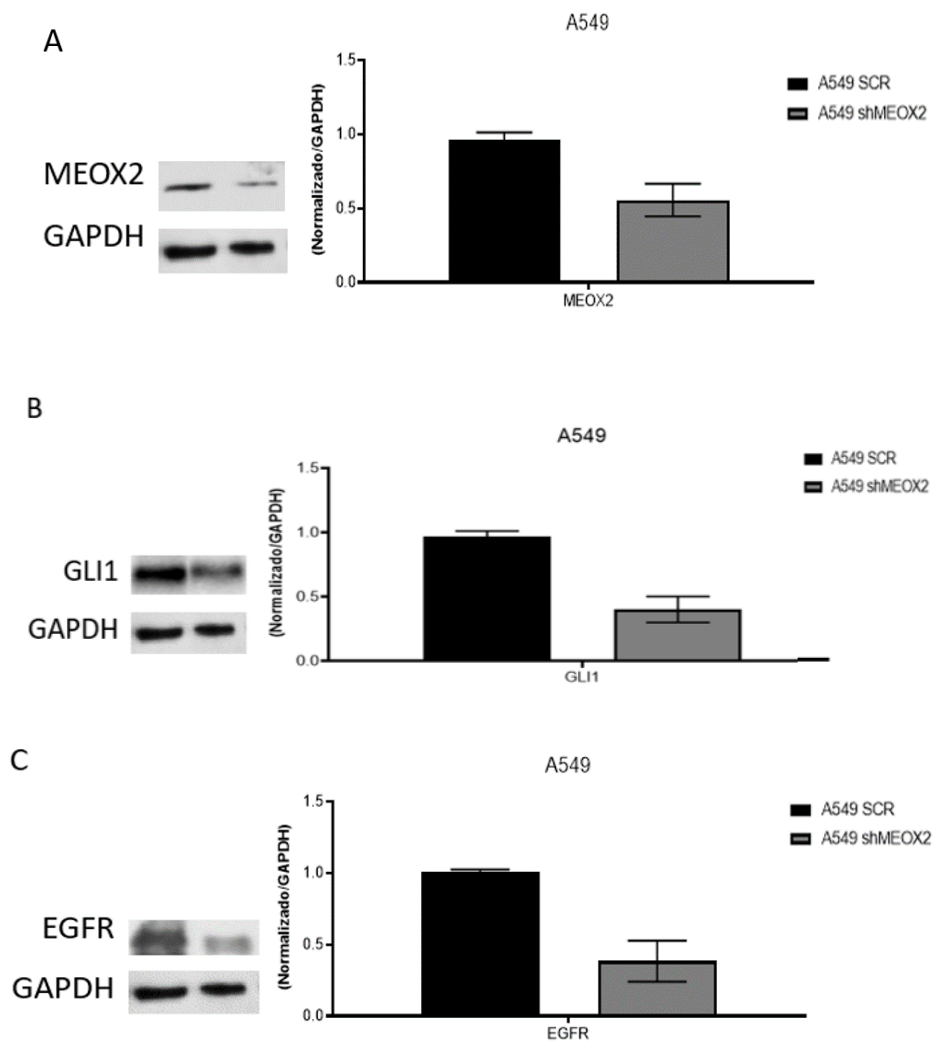
Los ensayos de validación se realizaron en la línea celular A549 de adenocarcinoma pulmonar con características morfológicas y bioquímicas de la célula alveolar pulmonar tipo II. En la figura 24, muestra la expresión relativa de mRNA para MEOX2 y sus genes blanco GLI1 y EGFR en la línea celular A549. La figura 24 A valida el silenciamiento de MEOX2 mediante short hairpin, reduciendo un 60% su expresión y se observó una tendencia a la disminución en los niveles de expresión tanto de GLI1 con un 10% (Figura 24 B) y EGFR con un 60% (Figura 24 C).



**Figura 24. Análisis de expresión de blancos génicos de MEOX2 en la línea celular A549 de cáncer de pulmón. A)** Validación del silenciamiento de MEOX2 **B)** Expresión de mRNA de GLI1 con el silenciamiento de MEOX2. **C)** Expresión de mRNA de EGFR con el silenciamiento de MEOX2. Se utilizó una secuencia *scrambled* SCR como control de transfección. El nivel relativo de expresión de proteína se normalizó con GAPDH, n=2.

Se determinaron los niveles de expresión de los genes blanco a nivel de proteína en la línea celular A549 para validar los resultados *in silico*. Con ello se identificó una tendencia a la disminución de la proteína de MEOX2 con su silenciamiento

(figura 25 A) y un patrón de disminución del 40 % sobre los niveles de proteína tanto para GLI1 (Figura 25 B) como para EGFR (Figura 25 C).



**Figura 25. Análisis de expresión de proteína de MEOX2, GLI1 y EGFR en la línea celular A549.** Se presenta una imagen representativa de los experimentos de Western Blot realizados, además se realizó un análisis cuantitativo para cada uno de ellos con 2 réplicas biológicas. **A)** Índice de cambio en expresión de proteína de MEOX2 con su silenciamiento. **B)** Índice de cambio en expresión de proteína de GLI1 con el silenciamiento de MEOX2. **C)** Índice de cambio en expresión de proteína de EGFR con el silenciamiento de MEOX2.

## 11. DISCUSIÓN

La inmunoprecipitación de cromatina, seguida de la secuenciación, ha sido una técnica experimental de relevancia para el estudio de las interacciones de los factores de transcripción en el DNA y de las modificaciones de histonas en la evaluación de la regulación epigenómica (Hokari *et al.*, 2020; Namani *et al.*, 2019), sin embargo, la identificación de tipos de interacciones representan un desafío bioinformático que ha requerido de una constante innovación en los modelos matemáticos y computacionales; al tiempo que de parámetros elaborados por el consorcio ENCODE (Landt *et al.*, 2012).

Los resultados masivos de ChIP-Seq son propensos a una serie de sesgos, derivados de los pasos del proceso experimental; tales como, la especificidad del anticuerpo, la inmunoprecipitación, la sensibilidad por baja afinidad con la proteína, la baja especificidad por la reactividad cruzada con proteínas no relacionadas y la fragmentación y/o el alineamiento.

Por lo anterior, se utilizan distintos filtros y controles durante la canalización del procesamiento de datos de ChIP-Seq para la obtención de resultados significativos con relevancia biológica (Meyer & Liu, 2014). Uno de los primeros es la evaluación de la calidad de las secuencias, en la que se determina la confiabilidad del uso de los datos y si requieren procesamientos posteriores; ésta primera evaluación incluye el recuento de la frecuencia relativa de los nucleótidos en cada posición de las lecturas para detectar posibles errores de acuerdo con las frecuencias esperadas; así mismo, evalúa la distribución de las puntuaciones de calidad de Phred para identificar lecturas con baja calidad global que pueden convertirse en problemas durante el análisis; también se mide la frecuencia de adaptadores y contaminantes de la muestra de DNA.

En ese sentido el reporte FastQC (figuras 7 y 8) mostró una calidad de secuenciación mayor a 20 (buena calidad), o de probabilidad  $\geq 99.0\%$  de corresponder a una llamada de base correcta, considerado estándar aceptable (de Sena Brandine & Smith, 2021). Otros parámetros importantes para obtener resultados con la menor cantidad de falsos positivos o negativos posibles en la

detección de sitios de unión *in vivo* para un regulador transcripcional son: la profundidad del muestreo en todo el genoma, la cobertura de secuenciación, la señal absoluta y la relación señal/ruido; para abordar este último se requiere estimar la distribución de fondo en cualquier posición genómica dada, por lo que los protocolos de ChIP-Seq sugieren secuenciar distintos controles, como el DNA-Input, donde no hay una Inmunoprecipitación asociada. Esto implica restar disponibilidad para la secuenciación de otras muestras problema, lo que también conlleva sumar costos. Este problema puede ser solucionado utilizando numerosos datos públicos disponibles (Awdeh *et al.*, 2021) que como se demostró es posible utilizar datos de ChIP-Seq externos tomando en cuenta las características de experimentación de la muestra problema.

Los resultados sobre el perfil de picos enriquecidos en todo el genoma, mostrados en la figura 9, arrojaron que MEOX2 se posiciona en todos los cromosomas incluyendo los sexuales; también fueron observadas regiones altamente densas o picos muy enriquecidos, lo que confirma su interacción a lo largo de todo el genoma.

Por otro lado, se observaron regiones sin señal en el inicio de los cromosomas 13, 14 y 15, indicando posibles zonas teloméricas o regiones centroméricas muy compactas en las intermedias sin señal, en los cromosomas 1, 9 y 16. Esto se explica debido a que la unión de los TF está influenciada por el estado de la cromatina, pues muchos se unen casi exclusivamente al estado abierto y tienen una afinidad por secuencias específicas; además de que existen TF que se unen al DNA como homo o heterodímeros, lo cual puede influir en el reconocimiento de sitios de unión diferentes (Jolma *et al.*, 2015).

MEOX2 pertenece a la familia de los genes HOX, de tipo Homeobox, localizados en distintos cromosomas; por ello se realizó un acercamiento al cromosoma 7 en el que se identificó un perfil de enriquecimiento a su largo, con uno notablemente enriquecido en su locus, pudiendo indicar su autorregulación; además de que, de acuerdo con distintas investigaciones los genes de tipo Homeobox pueden tener características funcionales, tanto promotoras como supresoras de tumores, según

el órgano o linaje celular específico donde se exprese (Joo *et al.*, 2016); tal como se demostró en HOXA9, regulado a la baja en los tejidos con cáncer de pulmón en comparación con los adyacentes no neoplásicos, y al alza en la leucemia linfocítica aguda (Rauch *et al.*, 2007).

La figura 11 reveló que el posicionamiento de MEOX2 se encuentra principalmente en regiones intergénicas y distales; esto muestra un primer acercamiento a la regulación de dicho factor de transcripción en regiones genómicas que se creían no funcionales, incluyendo aquellas altamente repetitivas. Este fenómeno también se observa en la figura 12 donde se visualiza la mayoría de la distribución de los picos de MEOX2 en regiones de 10 a 100 kb o mayores a 100; y a su vez, constituye la primera evidencia de una relación no descrita previamente para ese TF, respecto de su ocupación significativa en todo el epigenoma, en la línea celular A549 de cáncer de pulmón, más allá de lo reportado en 2017 en relación con los promotores génicos (Armas-López *et al.*, 2017). Esto sugiere una amplia participación de MEOX2 en la regulación de enhancers y superenhancers encontrados en elementos distales, lejos del sitio de inicio de la transcripción y dentro de los intrones o en genes próximos a los que regulan, de varios impulsores oncogénicos clave en distintas células tumorales, entre ellos NSCLC (Lovén *et al.*, 2013).

Diversos estudios han demostrado el rendimiento eficaz de distintos llamadores de picos, pero también diferencias en sus resultados; esto puede deberse a la configuración de sus algoritmos estadísticos, a la rigidez en sus parámetros, o al tipo de muestra analizada; por ejemplo, factores de transcripción versus el enriquecimiento de modificaciones de histonas de secuencias más largas (Jeon *et al.*, 2020).

En este sentido, el llamado de picos realizado con MACS2 (Figura 15) demostró ser más robusto que el de HOMER, por las ventajas del modelado empírico de “d” (distancia entre las cumbres de la cadena de Watson y Crick o Forward y Reverse) tras el comportamiento bimodal de las secuencias enriquecidas con ChIP. La importancia de este ajuste espacial en la construcción de los picos ha sido

demostrada al comparar resultados de análisis empíricos con distribuciones verdaderas de sitios motif conocidos, donde solo hubo variaciones de algunos pb (Zhang *et al.*, 2008).

Los resultados del análisis multi-comparativo demostraron que no hay impacto en el patrón de distribución de los picos en relación al TSS, ni en la anotación genómica; incluso cuando se mantuvo la condición de restado de fondo en HOMER y MACS2. Lo anterior, concuerda con la comparación y el uso de controles para el análisis de ChIP-Seq con datos públicos, donde se evaluaron distintas modificaciones de histonas, y en la que sus diferencias respondieron mayoritariamente al tipo de muestra analizado mostrando patrones distintos para aquellas marcas de cromatina abierta versus las marcas de cromatina cerrada o menos activas (Flensburg *et al.*, 2014). Sin embargo si se obtuvieron distintos números totales de picos con las 6 condiciones de restado de fondo, dicha variación impacto en el análisis de vías de señalización, evidenciando diferencias en los análisis funcionales, derivadas de la rigurosidad estadística y de las características de los DNA-Inputs; esto se observó en el input 1 y en el restado de fondo, también denominado sin input, al utilizar MACS2; mientras que, en el resto de las vías, el conjunto de genes no fue enriquecido en la misma proporción ni con la misma significancia.

Los algoritmos de llamada pico existentes, incluidos MACS así como algunos otros como CisGenome, PeakSeq, SPP y Sole-Search calculan la importancia del enriquecimiento en relación con el fondo local o combinan un umbral global con una comparación local, ya que se infiere que el fondo local es variable a lo largo del genoma pero reproducible entre experimentos repetidos. Sin embargo, se ha demostrado que los conjuntos de datos de DNA-Input varían entre réplicas técnicas y biológicas y que, por lo tanto, estos conjuntos de datos deben cumplir criterios muy estrictos. Aunque anteriormente no se podía prescindir del DNA-Input control actualmente la evidencia de los análisis comparativos de llamadores de picos han demostrado que el uso de conjuntos de datos de DNA-Input solo tiene ventajas limitadas ya que se ha observado que las llamadas máximas con

modelado de fondo algorítmico han demostrado ser igual de efectivas (de Boer *et al.*, 2014).

Los análisis de las vías de señalización confirmaron los resultados obtenidos por el grupo de investigación en 2017 del eje MEOX2-GLI1 en la vía de Sonic Hedgehog (Armas-López *et al.*, 2017), que demostró estar significativamente enriquecida con el análisis de ORA; eso, a su vez, se correlacionó con la resistencia funcional a las terapias de cisplatino y EGFR-TKI (Peralta-Arrieta *et al.*, 2022) resultantes de GSEA.

Con la finalidad de validar los resultados *in silico* se escogieron dos genes (GLI1 y EGFR) de las vías de señalización significativas y enriquecidas con los análisis ORA y GSEA. Los resultados mostraron que los niveles de expresión de mRNA de GLI1 disminuyeron con el silenciamiento de MEOX2, confirmando lo observado en 2017 sobre el posicionamiento de MEOX2 en el promotor de GLI1. Este fenómeno también se observó a nivel de proteína, lo que sugiere su posible correlación positiva.

GLI1 se considera el principal efector de activación de la vía Sonic Hedgehog, que se ha reportado al alza en distintos tumores de origen epitelial, entre ellos el cáncer de pulmón de células pequeñas; el de páncreas, colon y próstata, y los glioblastomas y meduloblastomas, demostrando que contribuye en la progresión del cáncer, la metástasis, la reparación de daños en el DNA y en la resistencia a la quimioterapia (Avery *et al.*, 2021; Shi *et al.*, 2010).



## 12. CONCLUSIONES

1. La calidad de las secuencias control DNA-input de datos públicos y la muestra experimental de CHIP de MEOX2, mostraron buena calidad con un alineamiento mayor al 95%.
2. Los picos enriquecidos en MEOX2 reflejaron su posicionamiento a lo largo de todo el genoma, principalmente en regiones distales intergénicas y mayor a 10 kb relativas al TSS.
3. El análisis comparativo determinó que el uso de controles DNA-Input da como resultado variabilidad en la determinación del número total de picos, sin embargo no modifica el perfil de distribución de estos.
4. El modelo computacional MACS2 genera un análisis más robusto al mejorar la reducción de la frecuencia de enriquecimiento de los picos, empero el uso de ambos llamadores de picos no modifico el porcentaje de distribución relativa al TSS ni la ubicación genómica de los picos.
5. El análisis bioinformático determinó 15 vías que se conservan entre los dos llamadores de picos siendo destacable la vía de Sonic Hedgehog y la de EGFR confirmando los estudios del grupo de investigación.
6. Los ensayos RT-qPCR y Western blot, determinaron una regulación positiva sobre GLI1 y EGFR tras el silenciamiento de MEOX2 en la línea celular A549, confirmando los análisis *in silico*.

### **13. PERSPECTIVAS**

1. Realizar análisis comparativos para la evaluación del uso de llamadores de picos ( MACS1, Cisgenome, Peak seq, SISSR)
2. Realizar análisis multi-comparativo sobre el uso de distintos alineadores y el uso de las versiones HG19 y HG38 como genomas de referencia en el análisis de ChIP-Seq.
3. Analizar la vía de señalización representativa del análisis bioinformático de adhesión focal por su relevancia en la oncología pulmonar en los procesos de invasión y metástasis.
4. Realizar análisis de los sitios motif conocidos de MEOX2.
5. Validar experimentalmente otros blancos génicos resultantes del análisis bioinformático.

## 15. BIBLIOGRAFÍA

1. Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B. A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., & Blankenberg, D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, *46*(W1), W537–W544. <https://doi.org/10.1093/nar/gky379>
2. Agents Classified by the IARC Monographs, Volumes 1–129 – IARC Monographs on the Identification of Carcinogenic Hazards to Humans. (s/f). Recuperado el 15 de mayo de 2022, de <https://monographs.iarc.who.int/agents-classified-by-the-iarc/>
3. Aggarwal, B. B., Shishodia, S., Sandur, S. K., Pandey, M. K., & Sethi, G. (2006). Inflammation and cancer: How hot is the link? *Biochemical Pharmacology*, *72*(11), 1605–1621. <https://doi.org/10.1016/j.bcp.2006.06.029>
4. Allis, C. D., Berger, S. L., Cote, J., Dent, S., Jenuwien, T., Kouzarides, T., Pillus, L., Reinberg, D., Shi, Y., Shiekhhattar, R., Shilatifard, A., Workman, J., & Zhang, Y. (2007). New Nomenclature for Chromatin-Modifying Enzymes. *Cell*, *131*(4), 633–636. <https://doi.org/10.1016/j.cell.2007.10.039>
5. Amos, C. I., Wu, X., Broderick, P., Gorlov, I. P., Gu, J., Eisen, T., Dong, Q., Zhang, Q., Gu, X., Vijaykrishnan, J., Sullivan, K., Matakidou, A., Wang, Y., Mills, G., Doheny, K., Tsai, Y.-Y., Chen, W. V., Shete, S., Spitz, M. R., & Houlston, R. S. (2008). Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nature Genetics*, *40*(5), 616–622. <https://doi.org/10.1038/ng.109>
6. Apte, R. S., Chen, D. S., & Ferrara, N. (2019). VEGF in Signaling and Disease: Beyond Discovery and Development. *Cell*, *176*(6), 1248–1264. <https://doi.org/10.1016/j.cell.2019.01.021>
7. Armas-López, L., Piña-Sánchez, P., Arrieta, O., de Alba, E. G., Ortiz-Quintero, B., Santillán-Doherty, P., Christiani, D. C., Zúñiga, J., & Ávila-Moreno, F. (2017). Epigenomic study identifies a novel mesenchyme homeobox2-GLI1 transcription axis involved in cancer drug resistance, overall survival and therapy prognosis in lung cancer patients. *Oncotarget*, *8*(40), 67056–67081. <https://doi.org/10.18632/oncotarget.17715>
8. Ávila-Moreno, F., Armas-López, L., Álvarez-Moran, A. M., López-Bujanda, Z., Ortiz-Quintero, B., Hidalgo-Miranda, A., Urrea-Ramírez, F., Rivera-Rosales, R. M., Vázquez-Manríquez, E., Peña-Mirabal, E., Morales-Gómez, J., Vázquez-Minero, J. C., Téllez-Becerra, J. L., Ramírez-Mendoza, R., Ávalos-Bracho, A., de Alba, E. G., Vázquez-Santillán, K., Maldonado-Lagunas, V., Santillán-Doherty, P., ... Zúñiga-Ramos, J. (2014). Overexpression of MEOX2 and TWIST1 Is Associated with H3K27me3 Levels and Determines Lung Cancer Chemoresistance and Prognosis. *PLoS ONE*, *9*(12), e114104. <https://doi.org/10.1371/journal.pone.0114104>
9. Ávila-Moreno, F., Armas-López, L., Álvarez-Moran, A. M., López-Bujanda, Z., Ortiz-Quintero, B., Hidalgo-Miranda, A., Urrea-Ramírez, F., Rivera-Rosales, R. M., Vázquez-Manríquez, E., Peña-Mirabal, E., Morales-Gómez, J., Vázquez-Minero, J. C., Téllez-

- Becerra, J. L., Ramírez-Mendoza, R., Ávalos-Bracho, A., de Alba, E. G., Vázquez-Santillán, K., Maldonado-Lagunas, V., Santillán-Doherty, P., ... Zúñiga-Ramos, J. (2016). Correction: Overexpression of MEOX2 and TWIST1 is Associated with H3K27me3 Levels and Determines Lung Cancer Chemoresistance and Prognosis. *PLoS ONE*, *11*(2), e0146569. <https://doi.org/10.1371/journal.pone.0146569>
10. Awdeh, A., Turcotte, M., & Perkins, T. J. (2021). WACS: Improving ChIP-seq peak calling by optimally weighting controls. *BMC Bioinformatics*, *22*, 69. <https://doi.org/10.1186/s12859-020-03927-2>
  11. Barlési, F., Gallegos-ruiz, M. I., Loundou, A., Span, S. W., Lefesvre, P., Kruyt, F. A. E., & Rodriguez, J. A. (2007). *Global histone modifications predict prognosis of resected non small-cell lung*.
  12. Bast, R. C. (Ed.). (2016). *Holland-Frei cancer medicine* (9th edition). John Wiley & Sons, Inc.
  13. Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., BRNAes, C. L., Bignell, H. R., Boutell, J. M., Bryant, J., Carter, R. J., Cheetham, R. K., Cox, A. J., Ellis, D. J., Flatbush, M. R., Gormley, N. A., Humphray, S. J., ... Smith, A. J. (2008). Accurate Whole Human Genome Sequencing using Reversible Terminator Chemistry. *Nature*, *456*(7218), 53–59. <https://doi.org/10.1038/nature07517>
  14. Biswas, S. K. (2015). Metabolic Reprogramming of Immune Cells in Cancer Progression. *Immunity*, *43*(3), 435–449. <https://doi.org/10.1016/j.immuni.2015.09.001>
  15. Böttcher, J. P., Bonavita, E., Chakravarty, P., Brees, H., Cabeza-Cabrerizo, M., Sammicheli, S., Rogers, N. C., Sahai, E., Zelenay, S., & Reis e Sousa, C. (2018). NK Cells Stimulate Recruitment of cDC1 into the Tumor Microenvironment Promoting Cancer Immune Control. *Cell*, *172*(5), 1022–1037.e14. <https://doi.org/10.1016/j.cell.2018.01.004>
  16. Brandt-Rauf, P. W. (1992). The molecular epidemiology of oncoproteins. *Scandinavian Journal of Work, Environment & Health*, *1*, 46–49.
  17. Brenner, D. R., McLaughlin, J. R., & Hung, R. J. (2011). Previous Lung Diseases and Lung Cancer Risk: A Systematic Review and Meta-Analysis. *PLoS ONE*, *6*(3), e17479. <https://doi.org/10.1371/journal.pone.0017479>
  18. Brotto, D. B., Siena, Á. D. D., de Barros, I. I., Carvalho, S. da C. e S., Muys, B. R., Goedert, L., Cardoso, C., Praça, J. R., Ramão, A., Squire, J. A., Araujo, L. F., & Silva, W. A. da. (2020). Contributions of HOX genes to cancer hallmarks: Enrichment pathway analysis and review. *Tumor Biology*, *42*(5), 101042832091805. <https://doi.org/10.1177/1010428320918050>
  19. Brzezińska, E., Dutkowska, A., & Antczak, A. (2013). The significance of epigenetic alterations in lung carcinogenesis. *Molecular Biology Reports*, *40*(1), 309–325. <https://doi.org/10.1007/s11033-012-2063-4>
  20. Cairns, R. A., Harris, I. S., & Mak, T. W. (2011). Regulation of cancer cell metabolism. *Nature Reviews Cancer*, *11*(2), 85–95. <https://doi.org/10.1038/nrc2981>
  21. Calcinotto, A., Kohli, J., Zagato, E., Pellegrini, L., Demaria, M., & Alimonti, A. (2019). Cellular Senescence: Aging, Cancer, and Injury. *Physiological Reviews*, *99*(2), 1047–1078. <https://doi.org/10.1152/physrev.00020.2018>

22. Calvayrac, O., Pradines, A., Pons, E., Mazières, J., & Guibert, N. (2017). Molecular biomarkers for lung adenocarcinoma. *European Respiratory Journal*, 49(4), 1601734. <https://doi.org/10.1183/13993003.01734-2016>
23. Cao, G., Huang, B., Liu, Z., Zhang, J., Xu, H., Xia, W., Li, J., Li, S., Chen, L., Ding, H., Zhao, Q., Fan, M., Shen, B., & Shao, N. (2010). Intronic miR-301 feedback regulates its host gene, *ska2*, in A549 cells by targeting MEOX2 to affect ERK/CREB pathways. *Biochemical and Biophysical Research Communications*, 396(4), 978–982. <https://doi.org/10.1016/j.bbrc.2010.05.037>
24. Chansky, K., Detterbeck, F. C., Nicholson, A. G., Rusch, V. W., Vallières, E., Groome, P., Kennedy, C., Krasnik, M., Peake, M., Shemanski, L., Bolejack, V., Crowley, J. J., Asamura, H., Rami-Porta, R., Goldstraw, P., Rami-Porta, R., Asamura, H., Ball, D., Beer, D. G., ... Yokoi, K. (2017). The IASLC Lung Cancer Staging Project: External Validation of the Revision of the TNM Stage Groupings in the Eighth Edition of the TNM Classification of Lung Cancer. *Journal of Thoracic Oncology*, 12(7), 1109–1121. <https://doi.org/10.1016/j.jtho.2017.04.011>
25. Chen, Y., Banda, M., Speyer, C. L., Smith, J. S., Rabson, A. B., & Gorski, D. H. (2010). Regulation of the Expression and Activity of the Antiangiogenic Homeobox Gene GAX/MEOX2 by ZEB2 and MicroRNA-221. *Molecular and Cellular Biology*, 30(15), 3902–3913. <https://doi.org/10.1128/MCB.01237-09>
26. Chen, Y., & Gorski, D. H. (2008). Regulation of angiogenesis through a microRNA (miR-130a) that down-regulates antiangiogenic homeobox genes GAX and HOXA5. *Blood*, 111(3), 1217–1226. <https://doi.org/10.1182/blood-2007-07-104133>
27. Chumakov, P. M. (2000). *Function of the p53 Gene: Choice between Life and Death*. 65(1), 13.
28. Collado, M., Blasco, M. A., & Serrano, M. (2007). Cellular Senescence in Cancer and Aging. *Cell*, 130(2), 223–233. <https://doi.org/10.1016/j.cell.2007.07.003>
29. Collins, L. G. (2007). Lung Cancer: Diagnosis and Management. *Lung Cancer*, 75(1), 8.
30. Corthay, A. (2014). Does the Immune System Naturally Protect Against Cancer? *Frontiers in Immunology*, 5, 197. <https://doi.org/10.3389/fimmu.2014.00197>
31. Coussens, L. M., & Werb, Z. (2002). Inflammation and cancer. *Nature*, 420(6917), 860–867. <https://doi.org/10.1038/nature01322>
32. Cuyàs, E., Verdura, S., Martin-Castillo, B., & Menendez, J. A. (2021). Metformin: Targeting the Metabolo-Epigenetic Link in Cancer Biology. *Frontiers in Oncology*, 10, 620641. Flensburg, C., Kinkel, S. A., Keniry, A., Blewitt, M. E., & Oshlack, A. (2014). A comparison of control samples for ChIP-seq of histone modifications. *Frontiers in Genetics*, 5. <https://www.frontiersin.org/article/10.3389/fgene.2014.00329>
33. Lee, R. (2020). *Gray's Anatomy (Fourth edition)*. Elsevier.
34. <https://doi.org/10.3389/fonc.2020.620641>
35. D'Aguanno, S., & Del Bufalo, D. (2020). Inhibition of Anti-Apoptotic Bcl-2 Proteins in Preclinical and Clinical Studies: Current Overview in Cancer. *Cells*, 9(5), 1287. <https://doi.org/10.3390/cells9051287>

36. de Sena Brandine, G., & Smith, A. D. (2021). Falco: High-speed FastQC emulation for quality control of sequencing data. *F1000Research*, 8, 1874. <https://doi.org/10.12688/f1000research.21142.2>
37. DeBerardinis, R. J., & Chandel, N. S. (2016). Fundamentals of cancer metabolism. *Science Advances*, 2(5), e1600200. <https://doi.org/10.1126/sciadv.1600200>
38. *Definición de cáncer—Diccionario de cáncer del NCI - Instituto Nacional del Cáncer* (nciglobal,ncienterprise). (2011, febrero 2). [NciAppModulePage]. <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/cancer>
39. Dick, F. A., & Rubin, S. M. (2013). Molecular mechanisms underlying RB protein function. *Nature reviews. Molecular cell biology*, 14(5), 297–306. <https://doi.org/10.1038/nrm3567>
40. Douville, J. M., Cheung, D. Y. C., Herbert, K. L., Moffatt, T., & Wigle, J. T. (2011). Mechanisms of MEOX1 and MEOX2 Regulation of the Cyclin Dependent Kinase Inhibitors p21CIP1/WAF1 and p16INK4a in Vascular Endothelial Cells. *PLoS ONE*, 6(12), e29099. <https://doi.org/10.1371/journal.pone.0029099>
41. Evan, G. I., & Vousden, K. H. (2001). Proliferation, cell cycle and apoptosis in cancer. *Nature*, 411(6835), 342–348. <https://doi.org/10.1038/35077213>
42. Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, et al. Global Cancer Observatory: Cancer Today. Lyon: International Agency for Research on Cancer; 2020 (<https://gco.iarc.fr/today>, consultado en febrero de 2022).
43. Füllgrabe, J., Kavanagh, E., & Joseph, B. (2011). Histone onco-modifications. *Oncogene*, 30(31), 3391–3403. <https://doi.org/10.1038/onc.2011.121>
44. *Goldstraw et al. - 2016—The IASLC Lung Cancer Staging Project Proposals f.pdf. (s/f)*. Recuperado el 16 de septiembre de 2021, de <https://www.jto.org/action/showPdf?pii=S1556-0864%2815%2900017-9>
45. Gowdy, K. M., & Fessler, M. B. (2014). *Emerging Roles for Cholesterol and Lipoproteins in Lung Disease*. 18.
46. Guengerich, F. P. (2006). Cytochrome P450s and other enzymes in drug metabolism and toxicity. *The AAPS Journal*, 8(1), E101–E111. <https://doi.org/10.1208/aapsj080112>
47. Guterres, A. N., & Villanueva, J. (2020). TARGETING TELOMERASE FOR CANCER THERAPY. *Oncogene*, 39(36), 5811–5824. <https://doi.org/10.1038/s41388-020-01405-w>
48. Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. *Cancer Discovery*, 12(1), 31–46. <https://doi.org/10.1158/2159-8290.CD-21-1059>
49. Hanahan, D., & Weinberg, R. A. (2000). The Hallmarks of Cancer. *Cell*, 100(1), 57–70. [https://doi.org/10.1016/S0092-8674\(00\)81683-9](https://doi.org/10.1016/S0092-8674(00)81683-9)
50. Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of Cancer: The Next Generation. *Cell*, 144(5), 646–674. <https://doi.org/10.1016/j.cell.2011.02.013>
51. Hanahan, D., & Weinberg, R. A. (2017). *Biological hallmarks of cancer*. 10.
52. Hanna, J. H., Saha, K., & Jaenisch, R. (2010). Pluripotency and Cellular Reprogramming: Facts, Hypotheses, Unresolved Issues. *Cell*, 143(4), 508–525. <https://doi.org/10.1016/j.cell.2010.10.008>

53. Hecht, S. S. (1999). Tobacco Smoke Carcinogens and Lung Cancer. *Journal of the National Cancer Institute*, 91(14), 17.
54. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H., & Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell*, 38(4), 576–589. <https://doi.org/10.1016/j.molcel.2010.05.004>
55. Herbst, R. S., Heymach, J. V., & Lippman, S. M. (2008). Lung Cancer. *New England Journal of Medicine*, 359(13), 1367–1380. <https://doi.org/10.1056/NEJMra0802714>
56. Hokari, S., Tamura, Y., Kaneda, A., Katsura, A., Morikawa, M., Murai, F., Ehata, S., Tsutsumi, S., Ishikawa, Y., Aburatani, H., Kikuchi, T., Miyazono, K., & Koinuma, D. (2020). Comparative analysis of TTF-1 binding DNA regions in small-cell lung cancer and non-small-cell lung cancer. *Molecular Oncology*, 14(2), 277–293. <https://doi.org/10.1002/1878-0261.12608>
57. Holland, P. W., Booth, H. A. F., & Bruford, E. A. (2007). Classification and nomenclature of all human homeobox genes. *BMC Biology*, 5(1), 47. <https://doi.org/10.1186/1741-7007-5-47>
58. Igney, F. H., & Krammer, P. H. (2002). Death and anti-death: Tumour resistance to apoptosis. *Nature Reviews Cancer*, 2(4), 277–288. <https://doi.org/10.1038/nrc776>
59. Illumina (2017) *An Introduction to Next-Generation Sequencing Technology*.
60. Jen, J., Tang, Y.-A., Lu, Y.-H., Lin, C.-C., Lai, W.-W., & Wang, Y.-C. (2017). Oct4 transcriptionally regulates the expression of long non-coding RNAs NEAT1 and MALAT1 to promote lung cancer progression. *Molecular Cancer*, 16, 104. <https://doi.org/10.1186/s12943-017-0674-z>
61. Kikuchi, S., Yamada, D., Fukami, T., Masuda, M., Sakurai-Yageta, M., Williams, Y. N., Maruyama, T., Asamura, H., Matsuno, Y., Onizuka, M., & Murakami, Y. (2005). Promoter Methylation of *DAL-1/4.1B* Predicts Poor Prognosis in Non-Small Cell Lung Cancer. *Clinical Cancer Research*, 11(8), 2954–2961. <https://doi.org/10.1158/1078-0432.CCR-04-2206>
62. Kim, H., Kwon, Y. M., Kim, J. S., Han, J., Shim, Y. M., Park, J., & Kim, D.-H. (2006). Elevated mRNA levels of DNA methyltransferase-1 as an independent prognostic factor in primary nonsmall cell lung cancer. *Cancer*, 107(5), 1042–1049. <https://doi.org/10.1002/cncr.22087>
63. Kim, J., & Bae, J.-S. (2016). Tumor-Associated Macrophages and Neutrophils in Tumor Microenvironment. *Mediators of Inflammation*, 2016, 6058147. <https://doi.org/10.1155/2016/6058147>
64. Kim, Y. J., Sheu, K. M., Tsoi, J., Abril-Rodriguez, G., Medina, E., Grasso, C. S., Torrejon, D. Y., Champhekar, A. S., Litchfield, K., Swanton, C., Speiser, D. E., Scumpia, P. O., Hoffmann, A., Graeber, T. G., Puig-Saus, C., & Ribas, A. (s/f). Melanoma dedifferentiation induced by IFN-γ epigenetic remodeling in response to anti-PD-1 therapy. *The Journal of Clinical Investigation*, 131(12), e145859. <https://doi.org/10.1172/JCI145859>
65. Krueger, F. (2022). *Trim Galore* [Perl]. <https://github.com/FelixKrueger/TrimGalore> (Original work published 2016)

66. Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T. R., & Weirauch, M. T. (2018). The Human Transcription Factors. *Cell*, *172*(4), 650–665. <https://doi.org/10.1016/j.cell.2018.01.029>
67. Landt, S. G., Marinov, G. K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B. E., Bickel, P., Brown, J. B., Cayting, P., Chen, Y., DeSalvo, G., Epstein, C., Fisher-Aylor, K. I., Euskirchen, G., Gerstein, M., Gertz, J., Hartemink, A. J., Hoffman, M. M., ... Snyder, M. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Research*, *22*(9), 1813–1831. <https://doi.org/10.1101/gr.136184.111>
68. Langevin, S. M., Kratzke, R. A., & Kelsey, K. T. (2015). Epigenetics of Lung Cancer. *Translational research : the journal of laboratory and clinical medicine*, *165*(1), 74–90. <https://doi.org/10.1016/j.trsl.2014.03.001>
69. Lemjabbar-Alaoui, H., Hassan, O. U., Yang, Y.-W., & Buchanan, P. (2015). Lung cancer: Biology and treatment options. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, *1856*(2), 189–210. <https://doi.org/10.1016/j.bbcan.2015.08.002>
70. Levy, S. E., & Boone, B. E. (2019). Next-Generation Sequencing Strategies. *Cold Spring Harbor Perspectives in Medicine*, *9*(7), a025791. <https://doi.org/10.1101/cshperspect.a025791>
71. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997 [q-bio]*. <http://arxiv.org/abs/1303.3997>
72. Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
73. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
74. Li, H.-P., Peng, C.-C., Chung, I.-C., Huang, M.-Y., Huang, S.-T., Chen, C.-C., Chang, K.-P., Hsu, C.-L., & Chang, Y.-S. (2013). Aberrantly hypermethylated Homeobox A2 derepresses metalloproteinase-9 through TBP and promotes invasion in Nasopharyngeal carcinoma. *Oncotarget*, *4*(11), 2154–2165.
75. Liang, K., & Keleş, S. (2012). Normalization of ChIP-seq data with control. *BMC Bioinformatics*, *13*(1), 199. <https://doi.org/10.1186/1471-2105-13-199>
76. Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z., & Zhang, B. (2019). WebGestalt 2019: Gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Research*, *47*(W1), W199–W205. <https://doi.org/10.1093/nar/gkz401>
77. Loomis, D., Huang, W., & Chen, G. (2014). The International Agency for Research on Cancer (IARC) evaluation of the carcinogenicity of outdoor air pollution: Focus on China. *Chinese Journal of Cancer*, *33*(4), 189–196. <https://doi.org/10.5732/cjc.014.10028>
78. Lugano, R., Ramachandran, M., & Dimberg, A. (2020). Tumor angiogenesis: Causes, consequences, challenges and opportunities. *Cellular and Molecular Life Sciences*, *77*(9), 1745–1770. <https://doi.org/10.1007/s00018-019-03351-7>



79. Malhotra, J., Malvezzi, M., Negri, E., Vecchia, C. L., & Boffetta, P. (2016). Risk factors for lung cancer worldwide. *European Respiratory Journal*, *48*(3), 889–902. <https://doi.org/10.1183/13993003.00359-2016>
80. Mantovani, F., Collavin, L., & Del Sal, G. (2019). Mutant p53 as a guardian of the cancer cell. *Cell Death & Differentiation*, *26*(2), 199–212. <https://doi.org/10.1038/s41418-018-0246-9>
81. Martin, G. S. (2003). Cell signaling and cancer. *CANCER CELL*, *8*.
82. Mehrotra, S., & Mitra, I. (2020). Origin of Genome Instability and Determinants of Mutational Landscape in Cancer Cells. *Genes*, *11*(9), 1101. <https://doi.org/10.3390/genes11091101>
83. Meyer, C. A., & Liu, X. S. (2014). Identifying and Mitigating Bias in Next-Generation Sequencing Methods for Chromatin Biology. *Nature reviews. Genetics*, *15*(11), 709–721. <https://doi.org/10.1038/nrg3788>
84. Moore, L. D., Le, T., & Fan, G. (2013). DNA Methylation and Its Basic Function. *Neuropsychopharmacology*, *38*(1), 23–38. <https://doi.org/10.1038/npp.2012.112>
85. Muenst, S., Läubli, H., Soysal, S. D., Zippelius, A., Tzankov, A., & Hoeller, S. (2016). The immune system and cancer evasion strategies: Therapeutic concepts. *Journal of Internal Medicine*, *279*(6), 541–562. <https://doi.org/10.1111/joim.12470>
86. Mundade, R., Ozer, H. G., Wei, H., Prabhu, L., & Lu, T. (2014). Role of ChIP-seq in the discovery of transcription factor binding sites, differential gene regulation mechanism, epigenetic marks and beyond. *Cell Cycle*, *13*(18), 2847–2852. <https://doi.org/10.4161/15384101.2014.949201>
87. Namani, A., Liu, K., Wang, S., Zhou, X., Liao, Y., Wang, H., Wang, X. J., & Tang, X. (2019). Genome-wide global identification of NRF2 binding sites in A549 non-small cell lung cancer cells by ChIP-Seq reveals NRF2 regulation of genes involved in focal adhesion pathways. *Aging*, *11*(24), 12600–12623. <https://doi.org/10.18632/aging.102590>
88. Nebbioso, A., Tambaro, F. P., Dell'Aversana, C., & Altucci, L. (2018). Cancer epigenetics: Moving forward. *PLoS Genetics*, *14*(6), e1007362. <https://doi.org/10.1371/journal.pgen.1007362>
89. Ohshima, J., Haruta, M., Arai, Y., Kasai, F., Fujiwara, Y., Ariga, T., Okita, H., Fukuzawa, M., Hata, J.-I., Horie, H., & Kaneko, Y. (2009). Two candidate tumor suppressor genes, MEOX2 and SOSTDC1, identified in a 7p21 homozygous deletion region in a Wilms tumor. *Genes, Chromosomes and Cancer*, *48*(12), 1037–1050. <https://doi.org/10.1002/gcc.20705>
90. Parikh, A. R. (2019). Lung Cancer Genomics. *Acta Medica Academica*, *48*(1), 78. <https://doi.org/10.5644/ama2006-124.244>
91. Park, P. J. (2009). ChIP-seq: Advantages and challenges of a maturing technology. *Nature Reviews Genetics*, *10*(10), 669–680. <https://doi.org/10.1038/nrg2641>
92. Pavlova, N. N., & Thompson, C. B. (2016). THE EMERGING HALLMARKS OF CANCER METABOLISM. *Cell metabolism*, *23*(1), 27–47. <https://doi.org/10.1016/j.cmet.2015.12.006>
93. Peralta-Arrieta, I., Trejo-Villegas, O. A., Armas-López, L., Ceja-Rangel, H. A., Ordóñez-Luna, M. del C., Pineda-Villegas, P., González-López, M. A., Ortiz-Quintero,

- B., Mendoza-Milla, C., Zatarain-Barrón, Z. L., Arrieta, O., Zúñiga, J., & Ávila-Moreno, F. (2022). Failure to EGFR-TKI-based therapy and tumoural progression are promoted by MEOX2/GLI1-mediated epigenetic regulation of EGFR in the human lung cancer. *European Journal of Cancer*, *160*, 189–205. <https://doi.org/10.1016/j.ejca.2021.10.032>
94. Perekatt, A. O., Shah, P. P., Cheung, S., Jariwala, N., Wu, A., Gandhi, V., Kumar, N., Feng, Q., Patel, N., Chen, L., Joshi, S., Zhou, A., Taketo, M. M., Xing, J., White, E., Gao, N., Gatzka, M. L., & Verzi, M. P. (2018). SMAD4 suppresses WNT-driven dedifferentiation and oncogenesis in the differentiated gut epithelium. *Cancer research*, *78*(17), 4878–4890. <https://doi.org/10.1158/0008-5472.CAN-18-0043>
  95. Pikor, L., Thu, K., Vucic, E., & Lam, W. (2013). The detection and implication of genome instability in cancer. *Cancer Metastasis Reviews*, *32*(3), 341–352. <https://doi.org/10.1007/s10555-013-9429-5>
  96. Rajabi, M., & Mousa, S. A. (2017). The Role of Angiogenesis in Cancer Treatment. *Biomedicines*, *5*(2), 34. <https://doi.org/10.3390/biomedicines5020034>
  97. Rauch, T. A., Zhong, X., Wu, X., Wang, M., Kernstine, K. H., Wang, Z., Riggs, A. D., & Pfeifer, G. P. (2008). High-resolution mapping of DNA hypermethylation and hypomethylation in lung cancer. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(1), 252–257. <https://doi.org/10.1073/pnas.0710735105>
  98. Rauch, T., Wang, Z., Zhang, X., Zhong, X., Wu, X., Lau, S. K., Kernstine, K. H., Riggs, A. D., & Pfeifer, G. P. (2007). Homeobox gene methylation in lung cancer studied by genome-wide analysis with a microarray-based methylated CpG island recovery assay. *Proceedings of the National Academy of Sciences*, *104*(13), 5527–5532. <https://doi.org/10.1073/pnas.0701059104>
  99. Reckamp, K. L. (Ed.). (2016). *Lung Cancer* (Vol. 170). Springer International Publishing. <https://doi.org/10.1007/978-3-319-40389-2>
  100. Risch, A., & Plass, C. (2008). Lung cancer epigenetics and genetics. *International Journal of Cancer*, *123*(1), 1–7. <https://doi.org/10.1002/ijc.23605>
  101. Saman, H., Raza, S. S., Uddin, S., & Rasul, K. (2020). Inducing Angiogenesis, a Key Step in Cancer Vascularization, and Treatment Approaches. *Cancers*, *12*(5), 1172. <https://doi.org/10.3390/cancers12051172>
  102. Seyfried, T. N., & Huysentruyt, L. C. (2013). On the Origin of Cancer Metastasis. *Critical reviews in oncogenesis*, *18*(1–2), 43–73.
  103. Sherr, C. J., & McCormick, F. (2002). The RB and p53 pathways in cancer. *Cancer Cell*, *2*(2), 103–112. [https://doi.org/10.1016/S1535-6108\(02\)00102-2](https://doi.org/10.1016/S1535-6108(02)00102-2)
  104. Sholl, L. M., Long, K. B., & Hornick, J. L. (2010). Sox2 Expression in Pulmonary Non-small Cell and Neuroendocrine Carcinomas. *Applied Immunohistochemistry & Molecular Morphology*, *18*(1), 55–61. <https://doi.org/10.1097/PAI.0b013e3181b16b88>
  105. Singh, N., Baby, D., Rajguru, J. P., Patil, P. B., Thakkannavar, S. S., & Pujari, V. B. (2019). Inflammation and Cancer. *Annals of African Medicine*, *18*(3), 121–126. [https://doi.org/10.4103/aam.aam\\_56\\_18](https://doi.org/10.4103/aam.aam_56_18)
  106. Soto, I., Grabowska, W. A., Onos, K. D., Graham, L. C., Jackson, H. M., Simeone, S. N., & Howell, G. R. (2016). Meox2 haploinsufficiency increases neuronal cell loss in

- a mouse model of Alzheimer's disease. *Neurobiology of aging*, 42, 50–60. <https://doi.org/10.1016/j.neurobiolaging.2016.02.025>
107. Spano, D., Heck, C., De Antonellis, P., Christofori, G., & Zollo, M. (2012). Molecular networks that regulate cancer metastasis. *Seminars in Cancer Biology*, 22(3), 234–249. <https://doi.org/10.1016/j.semcancer.2012.03.006>
  108. Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, 71(3), 209–249. <https://doi.org/10.3322/caac.21660>
  109. Suvà, M. L., Riggi, N., & Bernstein, B. E. (2013). Epigenetic Reprogramming in Cancer. *Science*, 339(6127), 1567–1570. <https://doi.org/10.1126/science.1230184>
  110. Tachon, G., Masliantsev, K., Rivet, P., Desette, A., Milin, S., Gueret, E., Wager, M., Karayan-Tapon, L., & Guichet, P.-O. (2021). MEOX2 Transcription Factor Is Involved in Survival and Adhesion of Glioma Stem-like Cells. *Cancers*, 13(23), 5943. <https://doi.org/10.3390/cancers13235943>
  111. Valcourt, U., Thuault, S., Pardali, K., Heldin, C.-H., & Moustakas, A. (2007). Functional role of Meox2 during the epithelial cyostatic response to TGF- $\beta$ . *Molecular Oncology*, 1(1), 55–71. <https://doi.org/10.1016/j.molonc.2007.02.002>
  112. van Zijl, F., Krupitza, G., & Mikulits, W. (2011). Initial steps of metastasis: Cell invasion and endothelial transmigration. *Mutation Research/Reviews in Mutation Research*, 728(1–2), 23–34. <https://doi.org/10.1016/j.mrrev.2011.05.002>
  113. Wadowska, K., Bil-Lula, I., Trembecki, Ł., & Śliwińska-Mossoń, M. (2020). Title: Genetic Markers in Lung Cancer Diagnosis: A Review. *International Journal of Molecular Sciences*, 21(13), 4569. <https://doi.org/10.3390/ijms21134569>
  114. Wang, Q., Zhan, Y., Pedersen, N. L., Fang, F., & Hägg, S. (2018). Telomere Length and All-Cause Mortality: A Meta-analysis. *Ageing Research Reviews*, 48, 11–20. <https://doi.org/10.1016/j.arr.2018.09.002>
  115. Welch, D. R., & Hurst, D. R. (2019). Defining the Hallmarks of Metastasis. *Cancer research*, 79(12), 3011–3027. <https://doi.org/10.1158/0008-5472.CAN-19-0458>
  116. Wilbanks, E. G., & Facciotti, M. T. (2010). Evaluation of Algorithm Performance in ChIP-Seq Peak Detection. *PLoS ONE*, 5(7), e11471. <https://doi.org/10.1371/journal.pone.0011471>
  117. Williams, A. B., & Schumacher, B. (2016). P53 in the DNA-Damage-Repair Process. *Cold Spring Harbor Perspectives in Medicine*, 6(5), a026070. <https://doi.org/10.1101/cshperspect.a026070>
  118. Witsch, E., Sela, M., & Yarden, Y. (2010). Roles for Growth Factors in Cancer Progression. *Physiology (Bethesda, Md.)*, 25(2), 85–101. <https://doi.org/10.1152/physiol.00045.2009>
  119. Wong, J. Y. Y., Zhang, H., Hsiung, C. A., Shiraishi, K., Yu, K., Matsuo, K., Wong, M. P., Hong, Y.-C., Wang, J., Seow, W. J., Wang, Z., Song, M., Kim, H. N., Chang, I.-S., Chatterjee, N., Hu, W., Wu, C., Mitsudomi, T., Zheng, W., ... Lan, Q. (2020). Tuberculosis infection and lung adenocarcinoma: Mendelian randomization and pathway analysis of genome-wide association study data from never-smoking Asian women. *Genomics*, 112(2), 1223–1232. <https://doi.org/10.1016/j.ygeno.2019.07.008>

120. Xu, W., Li, J., Li, L., Hou, T., Cai, X., Liu, T., Yang, X., Wei, H., Jiang, C., & Xiao, J. (2019). FOXD3 Suppresses Tumor-Initiating Features in Lung Cancer via Transcriptional Repression of WDR5. *STEM CELLS*, 37(5), 582–592. <https://doi.org/10.1002/stem.2984>
121. Yeung, K. T., & Yang, J. (2017). Epithelial–mesenchymal transition in tumor metastasis. *Molecular Oncology*, 11(1), 28–39. <https://doi.org/10.1002/1878-0261.12017>
122. Yu, G., Wang, L.-G., & He, Q.-Y. (2015). ChIPseeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*, 31(14), 2382–2383. <https://doi.org/10.1093/bioinformatics/btv145>
123. Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., & Liu, X. S. (2008). Model-based Analysis of ChIP-Seq (MACS). *Genome Biology*, 9(9), R137. <https://doi.org/10.1186/gb-2008-9-9-r137>Avery, J. T., Zhang, R., & Boohaker, R. J. (2021). GLI1: A Therapeutic Target for Cancer. *Frontiers in Oncology*, 11, 673154. <https://doi.org/10.3389/fonc.2021.673154>
124. Shi, T., Mazumdar, T., DeVecchio, J., Duan, Z.-H., Agyeman, A., Aziz, M., & Houghton, J. A. (2010). CDNA Microarray Gene Expression Profiling of Hedgehog Signaling Pathway Inhibition in Human Colon Cancer Cells. *PLoS ONE*, 5(10), e13054. <https://doi.org/10.1371/journal.pone.0013054>