



Instituto Tecnológico
de Buenos Aires



INSTITUTO LELOIR
FUNDACIÓN

PROYECTO FINAL DE CARRERA

**Análisis bioinformático para el tratamiento de
la heterogeneidad tumoral de cáncer
colorrectal con adenovirus oncolítico
direccionado transcripcionalmente con
promotores híbridos y sintéticos**

Autora:

Lourdes Maria HIRSCHSON ALVAREZ PRADO - 57220

Director:

Oswaldo PODHAJ CER

Índice general

1.	Resumen	7
2.	Agradecimientos	8
3.	Glosario de contracciones	9
Introducción		10
4.	Cáncer	10
4.1.	Generalidades del cáncer	10
4.2.	Causas del cáncer	11
4.3.	Heterogeneidad Tumoral	11
5.	Cáncer colorrectal	14
5.1.	Estadíos del CCR	14
5.2.	Tratamientos actuales del CCR	16
5.3.	Tratamientos actuales que consideran la heterogeneidad tumoral	19
5.3.1.	Anticuerpos monoclonales	20
5.3.2.	Inhibidores de puntos de control	20
5.3.3.	Inmunoterapias no específicas	20
5.3.4.	Terapia con células T	21
5.3.5.	Vacunas contra el cáncer	21
5.3.6.	Virus de replicación condicional	22
5.4.	Adenovirus	23
5.4.1.	Adenovirus oncolítico de replicación condicional	23
5.5.	Promotores	25
5.5.1.	A33	25
5.5.2.	AQP5	25
5.5.3.	CDC25A y CDC25B	26
5.5.4.	FSCN1	28

5.5.5.	GLI1	29
5.5.6.	NQO1	30
5.5.7.	VWA2	31
5.5.8.	Promotores híbridos	31
5.5.9.	Promotores sintéticos	32
6.	Bioinformática	35
6.1.	Minería de datos	36
6.2.	Minería de texto	36
6.3.	Bases de datos	36
6.3.1.	TRANSFAC	37
6.3.2.	JASPAR	39
6.3.3.	TRANSFAC vs JASPAR	40
7.	Hipótesis	41
8.	Objetivo General	42
9.	Objetivos Específicos	42
9.1.	Promotores híbridos	42
9.2.	Promotores sintéticos	42
	Materiales y Métodos	43
10.	MATERIALES	43
10.1.	Plataformas online	43
10.1.1.	ToppGene	43
10.1.2.	Gene Regulation - Match	45
10.1.3.	Promo	46
10.1.4.	Meme Suite- Fimo	47
10.1.5.	The Human Protein Atlas	48
10.1.6.	Gepia	49
10.1.7.	SnapGene	51
11.	MÉTODOS	51
11.1.	Búsqueda de los factores de transcripción	51
11.2.	Terminología	52
11.3.	Priorización génica	52
11.4.	Análisis de datos	52
11.4.1.	Factores de transcripción	52

11.4.2.	Posiciones	53
11.4.3.	Definición de variables para promotores híbridos	53
11.4.4.	Definición de variables para promotores sintéticos	54
11.4.5.	<i>Data Mining</i> para los factores de transcripción resultantes . . .	55
11.4.6.	Armado de los promotores híbridos con los respectivos factores de transcripción	55
11.4.7.	Selección de factores de transcripción para el armado de promotores sintéticos	56
Resultados		58
12.	Análisis de los factores de transcripción de los 8 promotores y las distancias al inicio de transcripción	58
12.1.	Primer análisis de los promotores	58
12.2.	Búsqueda de los factores de transcripción	60
12.3.	Terminología	60
12.4.	Posiciones	61
12.5.	Proceso de <i>Minería de texto</i> con los factores de transcripción determinados	61
12.6.	Resultado de los factores de transcripción de cada uno de los promotores y las distancias al inicio	62
12.7.	Resultado de la clasificación de los factores de transcripción según las variables definidas en el estudio	81
13.	Similitudes y diferencias entre los factores de transcripción de los 8 promotores .	115
14.	Promotores híbridos	120
15.	Promotores sintéticos	126
Discusión		129
16.	Situación actual de CCR	129
17.	Análisis de datos	130
18.	Factores de transcripción encontrados	131
19.	Perspectivas futuras	135
Conclusiones		136
Anexo		138
20.	Secuencia de promotores	138
Bibliografía		141

Índice de figuras

1.	Esquema representativo. Izquierda: tumor benigno. Derecha: tumor maligno.	11
2.	Esquema representativo de distintos tipos de heterogeneidad tumoral	12
3.	Heterogeneidad intratumoral y evolución clonal	14
4.	Capas del colon.	15
5.	Diagrama esquemático de la destrucción específica de células cancerosas de virus oncolíticos	22
6.	Estructura de adenovirus y regiones de transcripción	23
7.	Mecanismos de focalización tumoral por virus oncolíticos	24
8.	Expresión de la proteína AQP5 en CCR.	26
9.	Expresión de la proteína CDC25A en CCR.	27
10.	Expresión de la proteína CDC25B en CCR.	28
11.	Expresión de la proteína FSCN1 en CCR.	29
12.	Expresión de GLI1 en CCR	30
13.	Expresión de VWA2 en CCR	31
14.	Diseños típicos de promotores sintéticos para procariotas y eucariotas	34
15.	Estructura de la base de datos TRANSFAC	37
16.	Información obtenida en Gene Regulation comparando TRANSFAC Professional y TRANSFAC public.	38
17.	Formato de salida para una matriz asociada a un factor de transcripción.	40
18.	Selección de matrices de frecuencia de posición (PFM); grupo taxonómico: Vertebrados; <i>Single batch file (txt)</i> en formato MEME.	40
19.	Esquema del flujo de trabajo y la metodología en las aplicaciones de ToppGene Suite	44
20.	Esquema de la interfaz MEME Suite.	47
21.	Figura representativa de la cantidad de genes y anticuerpos registrados en <i>The Human Protein Atlas</i> en el correr de los años.	49
22.	Interfaz Gepia.	50

23.	Diagrama de flujo de la metodología empleada para los promotores híbridos. . .	56
24.	Diagrama de flujo de la metodología empleada para los promotores sintéticos. . .	57
25.	CDX2	117
26.	KLF5	118
27.	HNF4A	118
28.	HNF1A	118
29.	VDR	119
30.	NR1I2	119
31.	ELF3	120
32.	POU2F2.	120
33.	Estructura del promotor A33 con los factores identificados en la tabla 15. . . .	122
34.	Estructura del promotor AQP5 con los factores identificados en la tabla 15. . . .	122
35.	Estructura del promotor CDC25A con los factores identificados en la tabla 15. .	123
36.	Estructura del promotor CDC25B con los factores identificados en la tabla 15. .	123
37.	Estructura del promotor FSCN1 con los factores identificados en la tabla 15. . .	124
38.	Estructura del promotor GLI1 con los factores identificados en la tabla 15. . . .	124
39.	Estructura del promotor NQO1 con los factores identificados en la tabla 15. . . .	125
40.	Estructura del promotor VWA2 con los factores identificados en la tabla 15. . .	125
41.	Ejemplo de modelo de promotor híbrido de hasta 4 PETs.	125
42.	Resultados de Manika Sehgal et al., 2015, de 10 TF principales implicados en genes responsables de la expresión diferencial en CCR temprano.	126
43.	Primer promotor sintético	127
44.	Segundo promotor sintético	127
45.	Tercer promotor sintético	128
46.	Secuencia del promotor A33 de 410 pares de bases. En azul: +1	138
47.	Secuencia del promotor AQP5 de 600 pares de bases. En azul: +1	138
48.	Secuencia del promotor CDC25A de 450 pares de bases. En azul: +1	138
49.	Secuencia del promotor CDC25B de 450 pares de bases. En azul: +1	139
50.	Secuencia del promotor FASCIN de 600 pares de bases. En azul: +1	139
51.	Secuencia del promotor GLI1 de 604 pares de bases. En azul: +1	139
52.	Secuencia del promotor NQO1 de 840 pares de bases. En azul: +1	140
53.	Secuencia del promotor VWA2 de 632 pares de bases. En azul: +1	140

Índice de cuadros

1.	Sistema de clasificación TNM.	15
2.	Definición de variables para promotores híbridos.	53
3.	Definición de variables para promotores sintéticos.	54
4.	Resultados de estudios previos que asocian factores de transcripción a los promotores estudiados.	58
5.	Factores de transcripción para el promotor A33 con las posiciones correspondientes al inicio de transcripción y el valor-p ajustado de ToppGene.	62
6.	Factores de transcripción para el promotor AQP5 con las posiciones correspondientes al inicio de transcripción y el valor-p ajustado de ToppGene.	64
7.	Factores de transcripción para el promotor CDC25A con las posiciones correspondientes al inicio de transcripción y el valor-p ajustado de ToppGene.	68
8.	Factores de transcripción para el promotor CDC25B con las posiciones correspondientes al inicio de transcripción y el valor-p ajustado de ToppGene.	70
9.	Factores de transcripción para el promotor FSCN1 con las posiciones correspondientes al inicio de transcripción y el valor-p ajustado de ToppGene.	73
10.	Factores de transcripción para el promotor GLI1 con las posiciones correspondientes al inicio de transcripción y el valor-p ajustado de ToppGene.	77
11.	Factores de transcripción para el promotor NQO1 con las posiciones correspondientes al inicio de transcripción y el valor-p ajustado de ToppGene.	80
12.	Factores de transcripción para el promotor VWA2 con las posiciones correspondientes al inicio de transcripción.	81
13.	Resultado de todos los factores de transcripción identificados como estadísticamente significativos luego del ranqueo genético.	84
14.	Descripción de los factores de transcripción encontrados. Se describe el rol determinado en CCR (en caso de existir).	115
15.	Factores de transcripción de los ocho promotores en función de las variables <i>Específico de tejido</i> , <i>Específico de cáncer</i> o <i>Ambas</i>	116

1. Resumen

El principal problema que deriva en el fracaso de los tratamientos actuales de cáncer es la heterogeneidad tumoral, la cual genera que los clones resistentes a las terapias provoquen la recidiva de la enfermedad y su correspondiente metástasis. Es decir que los clones resistentes al tratamiento serán los que generen la recidiva. Esto es aplicable también a las nuevas terapias personalizadas/dirigidas que al tener un único blanco terapéutico van generando indirectamente que los clones que no sean atacados por la terapia escapen al tratamiento y sean los responsables de la recidiva.

El grupo de *Terapia Molecular y Celular* de la *Fundación Instituto Leloir (FIL)* ha generado adenovirus oncolíticos de replicación condicionada (AOL-RC) para el tratamiento del cáncer. Estos AOL-RC se multiplican exclusivamente en las células malignas ya que su replicación depende de un promotor genético obtenido de genes sobreexpresados en forma selectiva en el tumor. En forma similar a lo que se observa con las terapias personalizadas, la limitación del uso de un único promotor específico de tumor (PET), limita la utilidad de los virus solo a los clones malignos donde ese PET se encuentra activo. Los clones donde el PET no este activo no serán afectados por el virus y serán responsables de la recidiva tumoral.

Esta tesis tuvo como objetivo inicial generar PETs híbridos (PETH) con combinaciones de fragmentos de hasta 4 PETs diferentes partiendo de los PETs ya identificados por el laboratorio: A33, AQP5, CDC25A, CDC25B, FSCN1, GLI1, NQO1 y VWA2. Dichos genes regulados por los correspondientes promotores presentan una sobreexpresión en cáncer colorrectal (CCR) con respecto al tejido normal. Se realizó un análisis *in silico* de los fragmentos de interés de cada promotor para identificar los factores de transcripción (TF) y la especificidad de los mismos en cuanto a las siguientes variables: la enfermedad cáncer, la sobreexpresión en tejido intestinal, la relación con la actividad viral y la asociación con la respuesta inmunológica.

Empleando las plataformas online Match, Fimo, Promo y ToppGene determiné los TFs estadísticamente significativos de cada promotor asociados a las variables definidas. A través de herramientas de bioinformática de minería de datos y minería de texto seleccioné los TFs de interés de cada uno de los promotores. En esta tesis se brinda un análisis detallado de los TFs de cada promotor, explicando las similitudes y diferencias de expresión de cada una de las proteínas reguladas por dichos promotores en las células tumorales de colon y recto así como también la identificación de los principales TFs esenciales en la regulación de la actividad de cada promotor. El armado de promotores híbridos y sintéticos difiere considerablemente. Basado en el análisis de las secuencias de los promotores con sus TFs respectivos, se pueden identificar los fragmentos de cada promotor con mayor relevancia para el armado de los promotores híbridos. Por otro lado, para los promotores sintéticos, es necesario establecer la relevancia de cada TF en forma individual, la localización de los mismos dentro de cada promotor, la distancia entre ellos amén de otros parámetros que permitan finalmente predecir la mejor combinación posible de TFs para definir el promotor sintético.

2. Agradecimientos

A Osvaldo Podhajcer por la confianza y la oportunidad de realizar este trabajo en su laboratorio. A Eduardo Cafferata por el constante apoyo. A todos los becarios, David Gregorio Ríos, Julieta Barrera, Diana Aguilar Cortés y Mauro Heitrich por la continua motivación, optimismo y buena energía.

Menciones especiales a Daiana Ganiewich y Luciana Montivero, por su amabilidad y gran disposición. Me brindaron la ayuda que necesitaba, me guiaron en momentos que no sabía como proseguir, con una constante motivación y apoyo en el trabajo. Simplemente gracias!

A mis amigas de la facultad, con quienes nos fuimos acompañado y alentando en distintos momentos.

3. Glosario de contracciones

- 5FU: 5 Fluoruracilo
- AOL: adenovirus oncolítico
- AOL-RC: adenovirus oncolítico de replicación condicional
- CCR: Cáncer colorrectal
- CCSP-2: colon cancer secreted protein
- CG: cáncer gástrico
- CSC: células madre cancerosas
- EMT: Transición epitelial-mesenquimal
- PET: promotor específico de tumor
- PETH: promotor específico de tumor híbrido
- TF: Factor de transcripción
- TFs: Factores de transcripción
- TFBS: sitios de unión del factor de transcripción
- THPA: The Human Protein Atlas

Introducción

4. Cáncer

4.1. Generalidades del cáncer

El **cáncer** es la segunda causa de muerte en el mundo. Según las cifras presentadas por la Organización Mundial de la Salud (OMS) el 12 de septiembre de 2018, en 2015 la enfermedad ocasionó 8,8 millones de muertes. A su vez, la organización establece que cerca del 70 % de las muertes por cáncer se registran en países de ingresos medios y bajos. Según las cifras estimadas en 2015, los tipos de cáncer que causan mayor cantidad de muertes son:

- Pulmonar: 1,69 millones defunciones
- Hepático: 788 000 defunciones
- Colorrectal: 774 000 defunciones
- Gástrico: 754 000 defunciones
- Mamario: 571 000 defunciones

En particular, Argentina tiene una tasa de incidencia de 212 casos por 100.000 habitantes (sin distinguir entre sexos y considerando todos los tumores a excepción de piel no melanoma). Estas cifras provienen de las últimas estimaciones realizadas por la Agencia Internacional de Investigación sobre Cáncer (IARC) en base a datos disponibles a nivel mundial para el año 2018. Estas cifras implican que Argentina es un país con incidencia de cáncer media-alta (rango 177 a 245,6 por 100.000 habitantes), ubicándola a su vez en el séptimo lugar en Latinoamérica. Habiendo mencionado el estado general del país con respecto a la enfermedad [1], conocer los mecanismos de la misma es de vital importancia para así poder promover la detección temprana de la enfermedad y reducir la mortalidad.

El **cáncer** hace referencia a un grupo de más de 200 enfermedades que pueden afectar en cualquier parte del organismo. Si bien cada una de estas enfermedades es independiente, y por lo tanto, las características son individuales en cada caso, el denominador común de todas ellas es el mismo: las células cancerosas adquieren la capacidad de multiplicarse y diseminarse por todo el organismo sin control, pudiendo invadir partes adyacentes del cuerpo y propagarse a otros órganos. El término tumor hace referencia a una masa celular neoplásica, es decir, al crecimiento descontrolado y anormal del crecimiento de un tejido del organismo de manera

irreversible. Por un lado, el tumor benigno es aquel que se encuentra contenido y encapsulado en una región, mientras que el tumor maligno es aquel que tiene la capacidad de invadir a los tejidos vecinos, a través de los vasos sanguíneos. A esta capacidad se la conoce como metástasis. Estas masas celulares heterogéneas que crecen de manera autónoma invadiendo sin control se lo conoce como cáncer. Esto se visualiza en el esquema representativo de la figura 1.

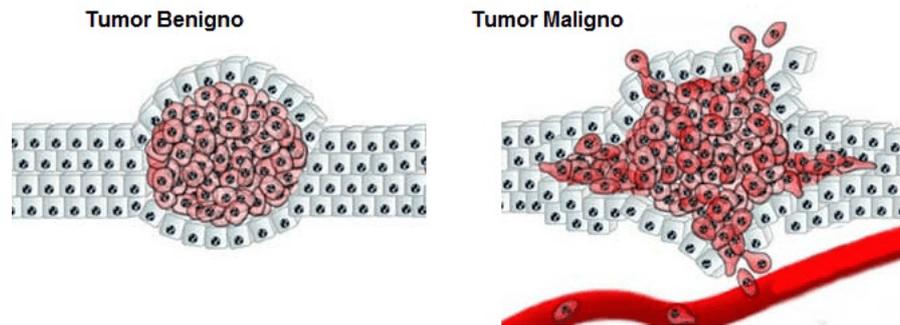


Figura 1: Esquema representativo. **Izquierda:** tumor benigno. **Derecha:** tumor maligno.

4.2. Causas del cáncer

El cáncer se produce por el proceso de transformación de las células normales en células tumorales. En los últimos años, diversos grupos de investigación se han avocado a identificar los procesos de mutaciones de las células que conducen a la proliferación de la enfermedad. La OMS sostiene que los factores genéticos del paciente así como ciertos factores externos, entre ellos las radiaciones ultravioletas y ionizantes, factores químicos como el amianto, los componentes del humo de tabaco, las aflatoxinas (contaminantes de los alimentos) y el arsénico (contaminante del agua de bebida), la edad o incluso ciertos virus, bacterias y parásitos, son elementales a la hora de explicar el por qué de la enfermedad en los individuos. La interacción entre estos factores puede terminar en mutaciones y pérdida de eficacia de los mecanismos de los procesos celulares [1].

4.3. Heterogeneidad Tumoral

En los últimos años, los estudios genómicos han demostrado el panorama complejo y heterogéneo del cáncer y sus posibles implicaciones para la respuesta al tratamiento y el pronóstico. Utilizando la secuenciación del genoma completo, estos estudios han demostrado que los tumores consisten en eventos somáticos, definidos por mutaciones y alteraciones del número de copias, que ocurren temprano en la evolución del tumor. Además, estos estudios han demostrado que en respuesta a las terapias contra el cáncer se genera la heterogeneidad espacial, es decir, una evolución ramificada de patrones mutacionales que pueden variar con el tiempo. Dada la existencia de dicha heterogeneidad en los tumores en la enfermedad metastásica avanzada, la eficacia de las terapias que tienen como target a las mutaciones somáticas, incluso combinadas con otras terapias, puede ser limitada en términos de lograr la cura de la enfermedad. La figura

2 del grupo de investigadores Mariam Jamal-Hanjani et al.; 2015 [2] esquematiza los distintos tipos de heterogeneidad genética en la biología del cáncer.

Por un lado, la más conocida es la **heterogeneidad tumoral interpaciente**, en la que no hay dos pacientes con el mismo subtipo de tumor que se comporten clínicamente de la misma manera, con o sin tratamiento. Esto puede estar relacionado con factores del huésped, como el microambiente tumoral y las variantes de la línea germinal que influyen en la respuesta al tratamiento, junto con las mutaciones somáticas únicas que pueden ocurrir dentro del tumor de cada paciente individual. La **heterogeneidad intratumoral** describe la existencia de distintas poblaciones celulares con características genéticas, epigenéticas y fenotípicas específicas dentro de los tumores y ha sido reconocida durante mucho tiempo. Las lesiones metastásicas en diferentes sitios secundarios pueden surgir de diferentes poblaciones celulares dentro de un tumor primario, lo que resulta en heterogeneidad entre metástasis, conocida como heterogeneidad intermetastásica. Además, dado que las lesiones metastásicas pueden adquirir nuevas mutaciones y evolucionar independientemente con cada división celular, también puede existir heterogeneidad dentro de una metástasis, conocida como **heterogeneidad intrametastásica**. Esto puede estar asociado con múltiples mecanismos de resistencia a los medicamentos adquiridos en el mismo paciente con enfermedad metastásica [2][3].

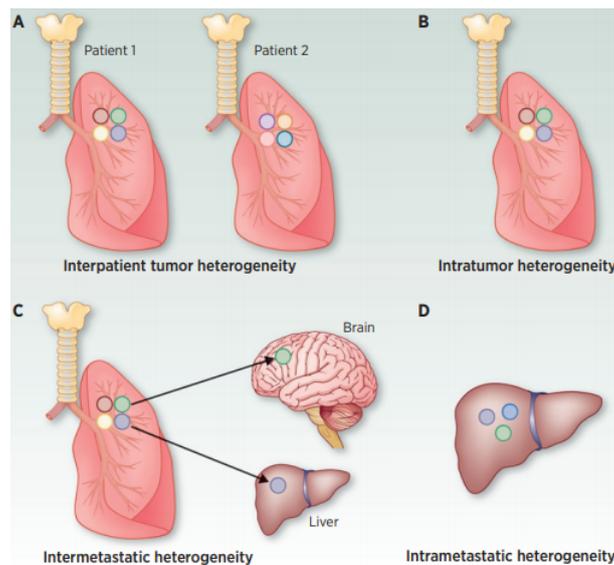


Figura 2: Diferentes tipos de heterogeneidad tumoral. **A**, heterogeneidad entre pacientes: la presencia de subclones únicos en el tumor de cada paciente. **B**, heterogeneidad intratumoral: la presencia de múltiples subclones dentro de un tumor primario que resulta en heterogeneidad entre las células tumorales. **C**, heterogeneidad intermetastásica: la presencia de diferentes subclones en diferentes lesiones metastásicas del mismo paciente; Algunos subclones pueden haberse derivado del tumor primario y algunos pueden haber surgido como resultado de alteraciones adquiridas dentro de cada lesión metastásica. **D**, heterogeneidad intrametastásica: la presencia de múltiples subclones dentro de una sola lesión metastásica. *Jamal-Hanjani et al., 2015*

La composición clonal y subclonal de cada tumor en el microambiente tumoral también puede influir en la evolución y progresión de los tumores. Las interacciones entre el tumor y las células del estroma, los cambios en el nivel de hipoxia o acidez, el aumento o la disminución del

infiltrado de células inflamatorias y la remodelación de la matriz extracelular pueden actuar como presiones de selección y conducir a una mayor heterogeneidad fenotípica, lo que puede influir potencialmente en la respuesta al tratamiento y, por lo tanto, en la evolución del tumor. Los tumores heterogéneos se componen de múltiples subclones y, bajo presiones de selección, como la quimioterapia, se pueden seleccionar ciertos subclones con resistencia intrínseca o adquirida, lo que permite que estos dominen una masa tumoral y potencialmente impulsen la progresión de la enfermedad. A través de múltiples eventos somáticos que afectan distintas vías de transducción de señales, un tumor puede desarrollar resistencia a los medicamentos dentro de la carga tumoral del mismo paciente, lo que plantea desafíos considerables para el diseño y la selección de combinaciones efectivas de medicamentos [2]. La figura 3 representa los diferentes subclones de los tumores primarios que pueden estar sujetos a diversas presiones de selección, que incluyen quimioterapia y factores microambientales como la hipoxia, el estroma infiltrante y las células inmunes, como mencioné anteriormente. Bajo la influencia de tales presiones de selección, los subclones con resistencia intrínseca (subclón verde) pueden superar una masa tumoral, lo que puede conducir a la progresión de la enfermedad, y / o los subclones pueden adquirir alteraciones somáticas (subclón púrpura), promoviendo la supervivencia celular, la proliferación y el tumor metastásico formación. El crecimiento de algunos subclones (subclón rojo) puede verse limitado por presiones de selección a las que son sensibles, por ejemplo, la terapia dirigida contra un subclón tumoral con una alteración somática sensible a la terapia.

Teniendo en cuenta que el sistema inmune tiene un rol fundamental en la progresión o en la cura de la enfermedad, la interacción entre el tumor y el sistema inmune del paciente fue muy estudiada a lo largo de los años. Los cánceres humanos albergan numerosas alteraciones genéticas y epigenéticas, generando neoantígenos potencialmente reconocibles por el sistema inmune. Sin embargo, aunque se observa una respuesta inmune endógena al cáncer en modelos preclínicos y pacientes, esta respuesta es ineficaz, porque los tumores desarrollan múltiples mecanismos de resistencia, incluidos los mecanismos locales de resistencia inmunosupresión y la inducción de tolerancia. Además, los tumores pueden explotar varias vías distintas para evadir activamente la destrucción inmune, incluidos los *puntos de control inmunológicos* endógenos que normalmente terminan las respuestas inmunes después de la activación del antígeno. Los microambientes tumorales pueden contrarrestar la respuesta inmune del huésped a través de mecanismos supresivos como la expresión de PDL-1 (programmed cell death ligand 1), o secreción de mediadores antiinflamatorios, entre otros [4] [5]. Los linfocitos infiltrantes de tumores pueden reconocer los neoantígenos presentados en la superficie de las células tumorales como no propios, promoviendo la activación mejorada de las células T y la infiltración del tumor de las células inmunes. Esto se visualiza como un ejemplo de respuesta en la figura 3 [2].

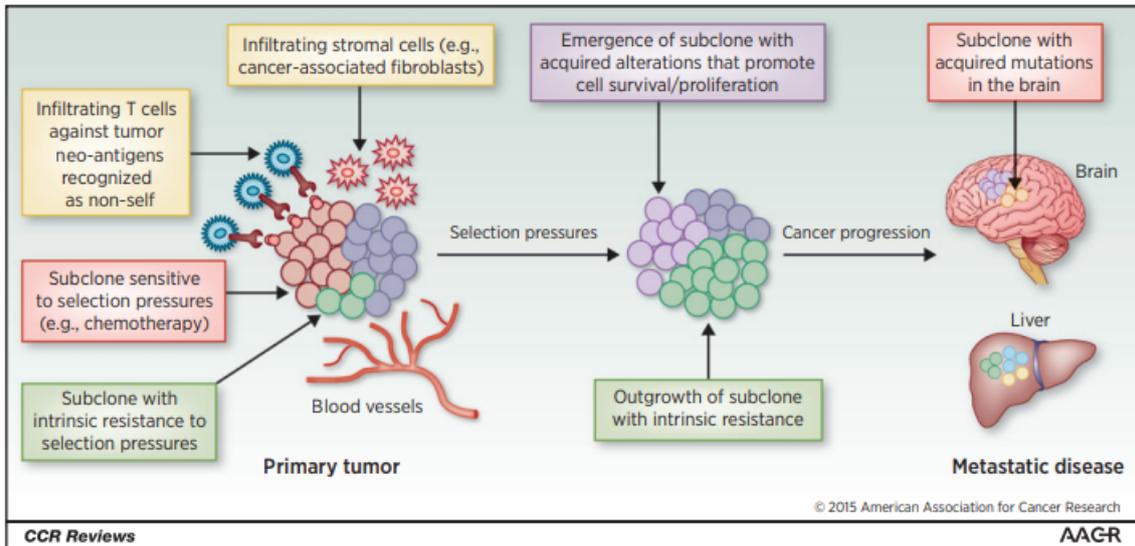


Figura 3: Heterogeneidad intratumoral y evolución clonal. *Jamal-Hanjani et al., 2015*

5. Cáncer colorrectal

El cáncer colorrectal (CCR) es el segundo cáncer más común en mujeres y el tercer cáncer más común en hombres [6].

En Argentina, según las estimaciones de incidencia del Observatorio Global de Cáncer de la OMS, el cáncer de mama es el de mayor magnitud en cuanto a ocurrencia del 2018: con un volumen de más de 21.000 casos al año, representa el 17% de todos los tumores malignos y casi un tercio de los cánceres femeninos. El segundo con mayor incidencia es el CCR, el cual para ambos sexos se estimaron 15.692 casos nuevos para el año 2018, concentrando el 13% del total de tumores [7].

5.1. Estadíos del CCR

La estadificación TNM se usa ampliamente y se acepta para estratificar a los pacientes en diferentes grupos pronósticos de la enfermedad. La tabla 1 hace esquematiza dicho sistema.

Sistema TNM	Descripción	Estadío
T	La extensión del tumor	T1: carcinoma in situ T2: tumor que invade la mucosa T3: tumor que invade la capa muscular T4: tumor que invade la serosa
N	La diseminación a los ganglios linfáticos	N0: ganglios negativos N1: de uno a tres ganglios afectados N2: más de tres ganglios afectados N3: ganglios afectados en troncos vasculares

M	Presencia de metástasis	M0: no hay metástasis M1: metástasis
---	-------------------------	---

Cuadro 1: Sistema de clasificación TNM.

La figura 4 esquematiza las capas del colon para un mayor entendimiento[8].

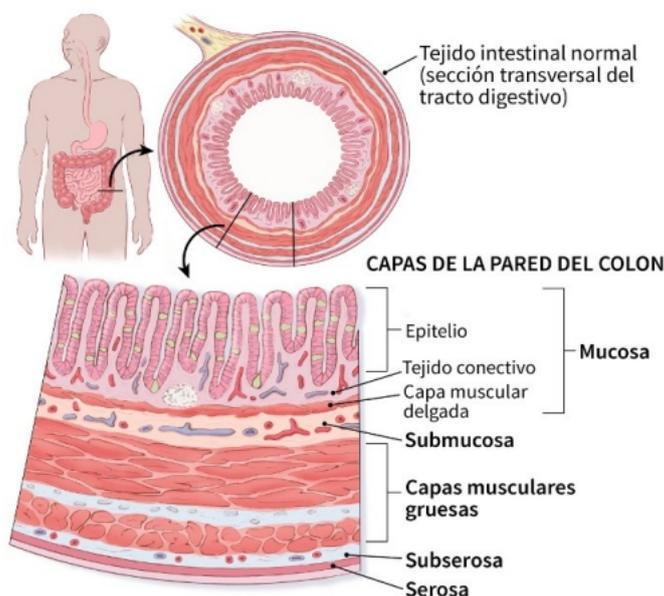


Figura 4: Capas del colon.

Esta clasificación es más detallada a medida que las letras TNM se acompañan con números, y a medida que el número crece, crece la agresividad del tumor[3]. A partir de esta clasificación, se establecen estadios generales I, II, III o IV:

- **0:** Carcinoma in situ: invasión confinada a la lámina propia de la mucosa del colon o del recto
- **I:** El tumor ha crecido traspasando varias capas (invasión de la submucosa (T1) o muscularis propia.(T2)), pero sin atravesar la capa muscular. Los ganglios no están afectados
- **II:** Invasión de todas las capas de la pared del colon o recto y en algunos casos se pueden haber afectado órganos vecinos. No están afectados los ganglios
 - IIa: Invasión a través de la muscularis propia (T3)
 - IIb: Invasión del peritoneo visceral (T4a)
 - IIc: Invasión de otros órganos o estructuras (T4b)
- **III:** Existen ganglios afectados, independientemente del tamaño del tumor

- IIIa:
 - T1 o T2 tumor con presencia de metástasis en 1–3 ganglios linfáticos regionales (N1)
 - Tumor con presencia de metástasis en 4–6 ganglios linfáticos regionales (N2a)
- IIIb:
 - Tumor T1 o T2 con presencia de metástasis en 7 o más ganglios linfáticos regionales (N2b)
 - T2 o T3 tumor con presencia de metástasis a N2a
 - T3 o T4a tumor con presencia de metástasis N1
- IIIc:
 - T3 o T4a tumor con presencia de metástasis a N2b
 - T4a tumor con presencia de metástasis de N2a
 - T4b tumor con presencia de metástasis N1 o N2
- **IV:** El tumor se ha diseminado y ha afectado a otros órganos a distancia (metástasis)
 - IVa: Presencia de metástasis confinada a un órgano sin metástasis peritoneales
 - IVb: Presencia de metástasis en más de un órgano
 - IVc: Presencia de metástasis al peritoneo con o sin afectación de otros órganos

5.2. Tratamientos actuales del CCR

Los tratamientos varían según la localización del carcinoma (recto o el resto del colon).

Cirugía: Es el principal tratamiento en los cánceres con etapas más tempranas.

- **Polipectomía y escisión local:** Los tumores en etapa 0 o I, es decir, en etapas iniciales, se busca extirpar los pólipos durante una colonoscopia. En este procedimiento se utiliza un tubo alargado y flexible con una pequeña videocámara en uno de sus extremos que se introduce por el recto para ser guiada hacia el interior del colon. Los siguientes procedimientos quirúrgicos pueden llevarse a cabo durante una colonoscopia:
 - En una polipectomía, el cáncer es extirpado como parte del pólipo.
 - Una escisión local implica un procedimiento ligeramente más complejo ya que se busca extirpar pequeños cánceres dentro del revestimiento del colon junto con una pequeña cantidad de tejido sano circundante sobre la pared del colon.
- **Colectomía:** es una cirugía que implica la extracción de parte o de todo el colon. También se extirpan los ganglios linfáticos cercanos.
 - Se conoce como **hemicolectomía**, **colectomía parcial** o **resección segmentaria** al procedimiento para extirpar solo una parte del colon. El cirujano extrae la parte del colon que contiene el cáncer junto con un pequeño segmento de colon normal en ambos lados. Se extirpa por lo general de un cuarto a un tercio del colon, dependiendo de la localización y el tamaño del cáncer. Luego se reconectan las secciones restantes

del colon. Al menos 12 ganglios linfáticos cercanos también son extraídos para que se puedan analizar y ver si hay cáncer.

- **Colectomía total** implica la extirpación de todo el colon. Solo se emplea si hay otro problema involucrado en el colon que no es cáncer, tales como enfermedad inflamatoria intestinal o pólipos.
- **Colostomía de derivación:** Los pacientes que tienen cáncer de colon y que el mismo se ha propagado a otras partes del cuerpo, se realiza una cirugía para aliviar la obstrucción sin extraer la parte del colon que contiene el cáncer. El colon se corta por encima del tumor y se adhiere a un estoma (una abertura en la piel del abdomen) para permitir que las heces fecales salgan del cuerpo. Es una cirugía contra la propagación del cáncer de colon. A menudo puede ayudar al paciente a recuperarse lo suficiente como para iniciar otros tratamientos (como la quimioterapia) [9].

Ablación: Cuando el CCR se ha propagado originando metástasis, a veces se pueden extraer con cirugía o eliminar mediante otras técnicas, como ablación o embolización. Si bien la técnica principal es la cirugía, la ablación se utiliza para destruir pequeños tumores entre 3 a 6 cm de diámetro sin tener que extirparlos, y esta técnica se emplea en los casos que los tumores vuelven a surgir tras la cirugía, para quienes el cáncer no se puede curar con la cirugía o para los que no puedan someterse a cirugía debido a otras razones. Esto podría ayudar a una persona a vivir más tiempo.

- **Radiación por radiofrecuencia:** utiliza ondas radiales de alta energía para destruir los tumores. Mediante una ecografía o tomografía por computadora, el médico coloca una sonda delgada parecida a una aguja a través de la piel y hacia el tumor. Luego se emite una corriente eléctrica hacia el extremo de la sonda, lo que libera ondas radiales de alta frecuencia que calientan el tumor destruyendo las células cancerosas.
- **Radiación por microondas:** Esta técnica de ablación más reciente se tiende a usar en el cáncer que se ha propagado al hígado. Los estudios por imágenes se usan para guiar la sonda con forma de aguja hacia el tumor. Microondas electromagnéticas se emiten a través de la sonda para generar altas temperaturas que matan los tumores con rapidez. Este tratamiento se ha usado contra tumores de mayor tamaño (de hasta 6 cm de amplitud).
- **Ablación con etanol:** También conocida como inyección percutánea de etanol, esta técnica consiste en inyectar alcohol concentrado directamente en el tumor para destruir las células cancerosas. Esto usualmente se hace a través de la piel usando una aguja que es guiada por ecografía o tomografías computarizadas.
- **Criocirugía:** La técnica consiste en destruir mediante congelación con una sonda de metal el tumor. A través de una ecografía, la sonda es guiada por la piel hasta alcanzar el tumor. Luego se pasan gases muy fríos a través del extremo de la sonda para congelar el tumor, lo que destruye las células cancerosas. Este método puede tratar tumores más grandes que las otras técnicas de ablación, pero a veces requiere de anestesia general. El tratamiento puede repetirse tantas veces como se requiera para eliminar todas las células cancerosas [10].

Tratamiento quimioterapéutico: La terapia sistémica es el uso de medicamentos para destruir las células cancerosas. Este tipo de medicamentos se administra a través del torrente

sanguíneo para llegar a las células cancerosas en todo el cuerpo. El profesional que receta terapias sistémicas es un oncólogo clínico, un médico que se especializa en el tratamiento del cáncer con medicamentos. En particular, la quimioterapia es el uso de fármacos con el fin de destruir las células cancerosas al poner fin a su capacidad para crecer y dividirse.

Un programa de quimioterapia, consiste en una cantidad específica de ciclos donde se administran en un período determinado. Un paciente puede recibir 1 fármaco por vez o una combinación de diferentes fármacos administrados al mismo tiempo.

La quimioterapia se puede administrar después de la cirugía para eliminar las células cancerosas restantes. En muchos casos, el tratamiento se utiliza en personas combinada con radioterapia antes de la cirugía para reducir el tamaño de un tumor y minimizar la posibilidad de que el cáncer reaparezca. La terapia se puede usar en distintas ocasiones durante el tratamiento contra CCR.

- La **quimioterapia adyuvante** se administra después de la cirugía. El objetivo es destruir algunas células cancerosas que podrían haber quedado durante la cirugía porque eran demasiado pequeñas como para verse, así como células cancerosas que podrían haber escapado del tumor principal y se establecieron en otras partes del cuerpo, siendo demasiado pequeñas como para verlas en estudios por imágenes. El fin de esto es disminuir la probabilidad de que el cáncer regrese.
- La **quimioterapia neoadyuvante** se administra antes de la cirugía para tratar de reducir el tamaño del cáncer y así hacer más fácil la cirugía. Se puede complementar con radioterapia. Este método se usa frecuentemente para tratar el cáncer de recto.
- Para los casos de cáncer en etapas avanzadas que se han extendido a otros órganos como el hígado, la quimioterapia puede utilizarse también para ayudar a encoger tumores y aliviar los síntomas. De esta manera, les permite a los pacientes mejorar la calidad de vida [11].

Con respecto a los medicamentos administrados, existen varios aprobados por la Administración de Alimentos y Medicamentos (Food and Drug Administration, FDA) para tratar el CCR. Los mecanismos de acción son los mismos, consisten en provocar una alteración celular, en la división celular, en síntesis de proteínas o en ácidos nucleicos, lo que lleva a una destrucción de las células en la división. Esto ocurre en células tanto tumorales como normales.

- Fluorouracilo (5-FU): es un fármaco específico del ciclo de fase celular, fase S, con actividad inhibitoria ya que interviene en la síntesis de ADN e inhibe en poco grado la formación de ARN. Promueve un desequilibrio metabólico que resulta en la muerte de la célula. Diversos estudios buscan estudiar los efectos de 5FU sobre no solo las células tumorales, también las respuestas inflamatorias y las medidas funcionales en un modelo de CCR en ratones y sugiere que los microbios intestinales pueden desempeñar un papel en algunas, pero no en todas, las perturbaciones relacionadas con 5FU [12].
- Capecitabina (Xeloda): el cual se administra oralmente (pastillas). Las ventajas que presenta el mismo es su capacidad de mimetizar la infusión continua de 5FU al tiempo que es un tratamiento cómodo con una mayor aceptación y cumplimiento por parte del enfermo,

evitando las complicaciones y costes económicos asociados a la administración intravenosa. A su vez, presenta una activación selectiva en el tumor que potencialmente posibilita un tratamiento localizado con la consiguiente mejora de la actividad antitumoral y reducción de la toxicidad sistémica [13].

- Irinotecán (Camptosar): fármaco empleado específicamente para cáncer de colon o de recto. Se administra de manera intravenosa. Interfiere en la acción de la enzima topoisomerasas I. Es un inhibidor de la topoisomerasa I, la cual controla la manipulación de la estructura del ADN necesaria para la replicación [14]. Si bien la quimioterapia basada en 5FU es el estándar de oro para el CCR, este medicamento ofrece pocos beneficios en ciertos estadios de CCR, y los regímenes basados en irinotecán puede una excelente alternativa [15].
- Oxaliplatino (Eloxatin): se identifica dicha droga con el fin de inducir la muerte inmunogénica¹ de las células de CCR, y este efecto determina su eficacia terapéutica en pacientes con CCR. [16]

5.3. Tratamientos actuales que consideran la heterogeneidad tumoral

A pesar de los grandes avances en los tratamientos actuales para el cáncer, la misma continúa siendo la segunda causa más frecuente de mortalidad, y en particular el CCR es la segunda causa de defunción dentro de los distintos cánceres en Argentina. Esto se debe a la falta de eficiencia en eliminar todas las células tumorales. Las terapias nombradas hasta el momento, tienden a disminuir el tamaño del tumor y eliminar ciertos clones del tumor. La combinación de terapias aumenta el índice de supervivencia de los distintos casos. Sin embargo, con frecuencia las células desarrollan mecanismos genéticos de resistencia que anulan la apoptosis, permitiendo a las mismas migrar y causar la recidiva de la enfermedad.

La **inmunoterapia** es un tipo de tratamiento del cáncer que ayuda al sistema inmunitario a combatir la enfermedad, estimulando así las defensas naturales del cuerpo. Este tipo de terapias involucra:

- Anticuerpos monoclonales
- Inhibidores de puntos de control
- Inmunoterapias no específicas
- Terapia con células T
- Vacunas contra el cáncer
- Terapia con virus oncolíticos

¹El concepto de muerte celular inmunogénica (MCI) incluye una muerte programada de las células cancerosas con la subsiguiente liberación de moléculas que dan una señal de peligro al sistema inmunológico.

5.3.1. Anticuerpos monoclonales

Son proteínas del sistema inmunitario creadas en el laboratorio con el propósito de unirse a blancos específicos en las células cancerosas. Se busca que los anticuerpos monoclonales reconozcan las células cancerosas y se unan a ellas, para luego ser destruidas por el sistema inmunitario. En particular, para CCR se utilizan los anticuerpos como **cetuximab** y **bevacizumab**. Cetuximab es un anticuerpo monoclonal anti-EGFR² que, utilizado solos o en combinación con quimioterápicos, permiten mejorar la respuesta al tratamiento en el CCR metastásico [17][18]. Por otro lado, bevacizumab, es un anticuerpo monoclonal humanizado dirigido al factor de crecimiento endotelial vascular, tiene un papel importante en el tratamiento del CCR metastásico, evitando la formación de nuevos vasos sanguíneos que irrigan al tumor y le permitan crecer mediante el suministro de nutrientes [19].

5.3.2. Inhibidores de puntos de control

Otros tipos de anticuerpos funcionan liberando los frenos del sistema inmunitario a fin de que este pueda destruir las células cancerosas. Las vías PD-1/PD-L1 y CTLA-4 son cruciales para la capacidad del sistema inmunitario de controlar el crecimiento del cáncer. Estas vías a menudo se denominan puntos de control inmunitarios (immune checkpoints). El sistema inmunitario responde al cáncer bloqueando estas vías con anticuerpos específicos denominados inhibidores de los puntos de control inmunitarios. Una vez que el sistema inmunitario es capaz de encontrar el cáncer y responder a este, puede detener o desacelerar el crecimiento del cáncer. En particular, la proteína 1 de muerte celular programada (PD1), limita las funciones efectoras de las células T dentro de los tejidos. Por lo tanto, al regular los ligandos para PD1, las células tumorales bloquean las respuestas inmunitarias antitumorales en el microambiente tumoral. Muchos tipos de cáncer utilizan estas vías para evadir el sistema inmunitario. A su vez, el antígeno 4 asociado a los linfocitos T citotóxicos (CTLA4), se encargan de modular la amplitud de la activación de las células T, haciendo que la respuesta inmune sea más ineficiente. La terapia anti-CTLA4 fue el primer agente en demostrar un beneficio de supervivencia en pacientes con melanoma avanzado y fue aprobada por la FDA en 2010 [20].

5.3.3. Inmunoterapias no específicas

Al igual que los anticuerpos monoclonales, las inmunoterapias no específicas también ayudan a que el sistema inmunitario destruya las células cancerosas. La mayoría de las inmunoterapias no específicas se administran después o al mismo tiempo que otros tratamientos para el cáncer, por ejemplo, la quimioterapia o la radioterapia. Sin embargo, algunas inmunoterapias no específicas se administran como el principal tratamiento del cáncer. Las terapias más utilizadas son:

²EGFR es una proteína transmembrana de la familia de los receptores de factores de crecimiento tipo tirosina cinasa. La activación de este receptor regula procesos como la proliferación, migración, invasión, angiogénesis y apoptosis. El EGFR desencadena dos vías de señalización principales: el eje RASRAF-MAPK involucrado primordialmente en la proliferación celular, y la vía PI3K-PTEN-AKT involucrada en la supervivencia celular, migración e invasión. La expresión aberrante del EGFR se ha asociado a diferentes tipos de cáncer epitelial, así como a un peor pronóstico y resistencia a la quimioterapia. Por lo tanto, se han desarrollado diferentes inhibidores de EGFR [17].

- **Interferones:** Los mismos ayudan al sistema inmunitario a combatir el cáncer y pueden desacelerar el crecimiento de las células cancerosas. Un tipo de interferón elaborado en un laboratorio se llama interferón alfa (Roferon-A [2a], Intron A [2b], Alferon [2a]). Este es el tipo de interferón que se usa con más frecuencia en el tratamiento del cáncer. Los efectos secundarios del tratamiento con interferón pueden incluir síntomas similares a los de gripe, un mayor riesgo de infecciones, erupciones cutáneas y cabello fino.
- **Interleuquinas:** Las interleuquinas ayudan al sistema inmunitario a producir células para combatir el cáncer. Una interleuquina elaborada en un laboratorio se llama interleuquina-2, IL-2, o aldesleuquina (Proleukin). Se utiliza para tratar principalmente el cáncer de riñón y el cáncer de piel, incluido el melanoma. Los efectos secundarios frecuentes del tratamiento con IL-2 incluyen aumento de peso y presión arterial baja. Algunas personas también pueden experimentar síntomas similares a los de la gripe [18].

5.3.4. Terapia con células T

La terapia con células T del receptor de antígeno quimérico (CAR) es una terapia celular que redirige las células T de un paciente para atacar y destruir específicamente las células tumorales. Los CAR son proteínas de fusión genéticamente modificadas compuestas de un dominio de reconocimiento de antígeno derivado de un anticuerpo monoclonal y una señalización intracelular y dominios coestimuladores de células T1, 2, 3, 4, 5. Teniendo en cuenta que el uso de las células T para terapia con CAR ha sido muy eficaz en el tratamiento de determinados cánceres de la sangre, este tipo de terapia aún está en proceso de estudio [21].

5.3.5. Vacunas contra el cáncer

Otro método existente es la vacuna contra el cáncer. Teniendo en cuenta el concepto de las vacunas, las cuales se administran a las personas sanas para ayudar a prevenir infecciones, utilizando gérmenes debilitados o muertos como virus o bacterias para iniciar una respuesta inmunitaria en el organismo, se busca preparar el sistema inmunitario para defenderse de estos gérmenes. De esta manera, se evita que las personas contraigan las infecciones. Lo mismo se aplica a las enfermedades cancerígenas. Una vacuna expone al sistema inmunitario a un antígeno, buscando de esta manera que el sistema inmunitario de la persona ataque a las células cancerosas y destruya ese antígeno o los materiales relacionados. Existen 2 tipos de vacunas contra el cáncer: vacunas para prevención y vacunas para tratamiento [18].

Por un lado, las de prevención tienen como blanco a los virus que pueden causar ciertos tipos de cáncer. Pueden ayudar a proteger contra algunos tipos de cáncer, pero no se dirigen directamente a las células cancerosas ya que aún no se han formado o encontrado dichas células.

Las vacunas para el tratamiento del cáncer son diferentes de las vacunas que funcionan contra los virus, como las preventivas, ya que buscan preparar al sistema inmunitario para que lance un ataque contra las células cancerosas en el organismo. Estas vacunas tienen el objetivo de lograr que el sistema inmunitario ataque a una enfermedad que ya existe. Algunas vacunas para el tratamiento del cáncer se componen de células cancerosas, partes de células o antígenos puros,

es decir, ciertas proteínas en las células cancerosas. A veces, las vacunas contienen las propias células inmunitarias extraídas de un paciente y expuestas a las sustancias en el laboratorio para su proliferación. Una vez que la vacuna está lista, se inyecta en el cuerpo para aumentar la respuesta inmunitaria contra las células cancerosas [22].

5.3.6. Virus de replicación condicional

Para explicar esta terapia primero se debe mencionar el concepto de los virus oncolíticos. Los virus oncolíticos son virus utilizados para identificar y atacar tumores que ya se han formado, buscando que el mismo se propague selectivamente en el tejido tumoral y lo destruye sin causar un daño excesivo a los tejidos normales no cancerosos. [23]. Una vez entregado a una célula diana susceptible, el genoma viral usurpa la maquinaria biosintética celular para reproducirse, permitiendo que el mismo se propague a las células adyacentes, lo que conduce a un patrón característico de destrucción de tejidos. De esta manera, la lisis de estas células desencadenan una respuesta inflamatoria, despertando las respuestas inmunes innatas y adaptativas del tumor [24]. Para que la terapia sea exitosa, los virus oncolíticos deben ser capaces de penetrar las defensas del huésped para acceder a los tumores en crecimiento, ya sean primarios o metastásicos. También deben ser capaces de propagarse lo suficiente en el sitio objetivo para destruir los tumores infectados antes de que el sistema inmunitario controle y elimine la infección. La figura 5 es un esquema representativo de la actividad viral de un virus oncolítico.

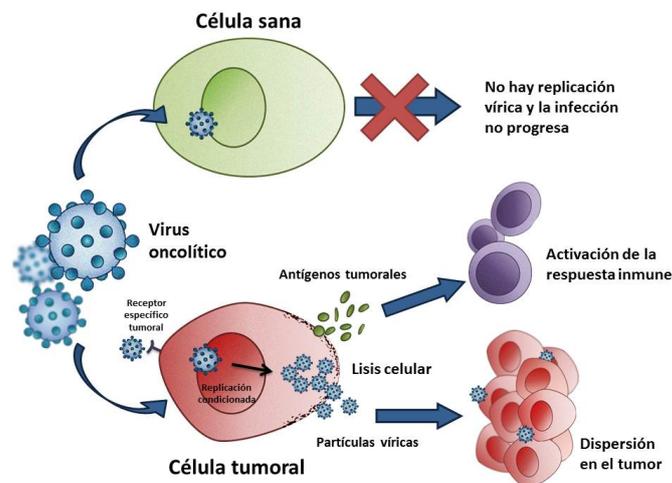


Figura 5: Diagrama esquemático de la destrucción específica de células cancerosas de virus oncolítico. El mismo puede replicarse y destruir las células cancerosas por oncólisis específica del cáncer mientras evita las células normales (*Joung-Woo Choi et al., 2015[25]*).

Por seguridad, los virus oncolíticos deben ser altamente específicos para el cáncer y al tejido de interés para así causar daños mínimos en la destrucción a los tejidos normales. En esta tesis de grado se busca realizar un análisis bioinformático para lograr una modificación genética de los virus que aumente la especificidad y sensibilidad del mismo. En particular, el virus de interés en este proyecto es el **adenovirus**.

5.4. Adenovirus

Los adenovirus son virus sin envoltura con un ADN lineal de doble cadena, de aproximadamente 36 kb. Hay 7 especies de adenovirus humanos (A a G) y 57 serotipos. Pertenecen a la familia *Adenoviridae*, y los diferentes serotipos están asociados con distintas enfermedades. En los seres humanos, pueden generar graves infecciones en las vías respiratorias, cistitis hemorrágica, gastroenteritis y conjuntivitis. El genoma viral consiste en los genes tempranos (E1, E2, E3 y E7) y los genes tardíos (L1, L2, L3, L4 y L5). La figura 6 esquematiza la secuencia del adenovirus.

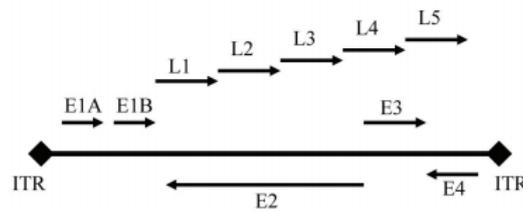


Figura 6: Estructura de adenovirus y regiones de transcripción. El genoma lineal adenoviral de ADN bicatenario está unido covalentemente en sus extremos a proteínas terminales a través de repeticiones terminales invertidas (ITR). La primera unidad de transcripción viral que se expresa es E1A que activa las otras transcripciones de adenovirus tempranas (E1B, E2, E3 y E4). La transcripción génica tardía (L1-L5) se activa después del inicio de la replicación del ADN. Las flechas indican la dirección de la transcripción. (*Bart Thaci et al., 2011[27]*).

Después de la infección de la célula objetivo y el desmontaje del virus en el endosoma temprano, el ADN viral ingresa al núcleo a través de los poros con la ayuda de microtúbulos y dineína. La primera unidad que se transcribe al ingresar el genoma viral es E1A. Esta proteína actúa como factor activador de otras unidades transcripcionales, controlando el ciclo celular y la expresión de los otros genes tempranos. En la fase S el virus toma control de la maquinaria celular para poder replicar así su propio genoma. Su ausencia impide la replicación viral. Además, debido a la unión entre E1A y retinoblastoma (Rb), se permite la liberación de E2F, lo que produce la progresión del ciclo celular [26] [27].

5.4.1. Adenovirus oncolítico de replicación condicional

Los adenovirus oncolítico de replicación condicional (AOL-RC) son virus con selectividad tumoral generada mediante ingeniería genética que replican y lisan selectivamente a las células tumorales. Una de las maneras de realizar esto es mediante la **regulación transcripcional** de los genes virales [28].

Como bien se mencionó anteriormente, uno de los mayores desafíos que enfrenta la terapia contra el cáncer es generar estrategias de tratamiento específicas para el tumor. La mayoría de los tumores malignos metastásicos y localmente avanzados son refractarios al tratamiento. La terapia génica espera lograr esto dirigiendo la actividad de los genes terapéuticos específicamente a los sitios de la enfermedad.

Se conoce como **promotores** a las secuencias de ADN que controlan la iniciación de la transcripción de una determinada porción del ADN a ARN. Por lo tanto, contienen la información necesaria para activar o desactivar el gen que regula. Su función es mediar y controlar el inicio de transcripción del gen que se encuentra situado río abajo de dicho promotor. La terapia génica para el cáncer requiere que la expresión de determinado gen sea específico y restringido al sitio tumoral de interés. Por lo tanto, para poder lograr una terapia selectiva con altos niveles de transcripción en poblaciones celulares definidas se deben tener promotores celulares o elementos reguladores específicos. Mientras mayor sea el nivel de especificidad de expresión de la proteína, mayor será el interés del promotor de la misma ya que permite que la orientación de la terapia sea más exacta al sitio de interés. La figura 7 esquematiza como se puede lograr la orientación transcripcional del virus oncolítico con promotores.

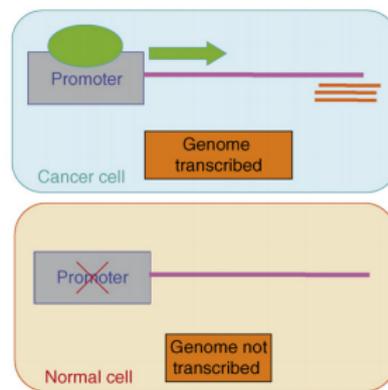


Figura 7: Mecanismos de focalización tumoral por virus oncolíticos. Orientación transcripcional. Un gen viral esencial se coloca bajo el control de un promotor específico de tumor (algunos promotores de virus son naturalmente específicos de tumor). Típicamente, el gen seleccionado codifica una proteína viral temprana que es esencial para completar con éxito el ciclo de vida del virus. En el caso del adenovirus, este gen es E1A (*Stephen J Rusell et al., 2007[24]*).

Muchos **promotores específicos de tejido** han sido aislados y caracterizados, y han sido utilizados en modelos experimentales de terapia génica contra el cáncer. Teniendo en cuenta que los promotores específicos de tejido también son activos en el tejido normal del que se originó el tumor, se deben considerar mecanismos de especificidad adicionales para que el vector resultante sea útil para el suministro sistémico. Por lo tanto, se busca introducir un nivel adicional de especificidad en el promotor específico del tejido, como una especificidad para las células en proliferación o una selectividad para otras afecciones que son características de las células tumorales. Se conoce como **promotores específicos de tumor (PET)** a aquellos cuyos genes regulados han sido identificados con actividad en condiciones patológicas, sobre-expresado en células cancerígenas y no en células normales [28]. Por lo tanto, en el caso de los adenovirus oncolíticos se emplean estos PET para dirigir la terapia únicamente a las células target. Teniendo en cuenta que E1A es el primer gen transcripto del adenovirus, se busca modificar genéticamente la secuencia del mismo, para reemplazar el promotor natural de E1A por un PET. De esta manera, se puede dirigir la replicación selectiva del virus a un tejido tumoral específico [29]. Es de suma importancia para una terapia con alta eficacia generar un adenovirus oncolítico con capacidad de replicación tumoral restringida al tejido colorrectal tumoral. En el caso de CCR, los promotores seleccionados por el *Instituto Leloir*, en particular

el laboratorio dirigido por *Oswaldo Podhajcer*, son: A33, AQP5, CDC25A, CDC25B, FSCN1, GLI1, NQO1 y VWA2.

5.5. Promotores

5.5.1. A33

El antígeno A33 es una proteína unida a la membrana expresada casi exclusivamente por las células epiteliales intestinales[30]. El gen A33 se encuentra en el cromosoma 1 de los humanos. El nivel de su expresión es consistente en todo el eje rostrocaudal del intestino humano y del ratón. En el colon, se encuentra una fuerte expresión en las membranas basolaterales de las células proliferantes en las regiones inferiores de las criptas y en las células diferenciadoras en las regiones superiores de las criptas, así como en las células epiteliales de las criptas en el intestino delgado [31]. Las pruebas en muestras tumorales demostraron que solo los tumores del tracto gastrointestinal son consistentemente positivos para A33. A33 se encuentra en el 95 % de los cánceres colorrectales primarios y metastásicos, con una expresión uniforme en todos los tumores en la mayoría de los casos. A33 también se detecta en el 63 % de los cánceres gástricos, con expresión uniforme en el 45 % de los casos. El ochenta y tres por ciento de los cánceres gástricos de tipo intestinal fueron positivos para A33, y aproximadamente el 50 % de los cánceres de tipo difuso y mucinoso fueron positivos para A33. A33 se expresó en el 50 % de los cánceres pancreáticos pero con marcada heterogeneidad. Otros cánceres epiteliales, sarcomas, tumores neuroectodérmicos y neoplasias linfoides fueron generalmente A33 negativos. Por lo tanto, A33 es el primer ejemplo de un antígeno de membrana epitelial específico de órgano expresado constitutivamente que permite la focalización de tumores altamente específicos en pacientes con cáncer gastrointestinal [32]. Estos resultados coinciden con los de Zebin Mao, et al. (2003) y Cameron N. Johnstone et al. (2002), los cuales muestran cómo la sobreexpresión de A33 está estrictamente asociada a las células de cáncer colorrectal. Por lo tanto, la evidencia acumulada valida A33 como un marcador definitivo para las células de CCR, para tumores primarios y metástasis [31][30].

5.5.2. AQP5

Las aquaporinas (AQP) son una familia de proteínas integrales, principalmente canales de agua de membrana, necesarios en el transporte de agua a través de muchos epitelios secretores y absorbentes. Se encontraron algunos subtipos de AQP involucrados en el transporte de otras moléculas, como el glicerol y la urea. Se ha revelado una expresión alterada de AQP en varios tipos de tumores en función de su localización tisular específica. En particular, AQP5 es un gen que pertenece al cromosoma 12 humano. Está ampliamente sobreexpresado en el cáncer de páncreas y parece estar involucrado en la proliferación celular. En el CCR, la sobreexpresión de AQP5 se asocia con metástasis, lo que sugiere que la proteína es importante en la progresión del cáncer [33][34]. Dado que el crecimiento, el desarrollo, la invasión y las metástasis tumorales dependen en gran medida del microambiente tumoral y su metabolismo, la cantidad de moléculas de agua juega un papel importante en la progresión de los tumores epiteliales malignos, lo que explica la sobreexpresión de AQP5 en el CCR[35]. La figura 8 representa la sobre-expresión de la proteína en CCR comparado con el tejido normal. Tanto Wang et al., 2018, como Zhikuan

He et al., 2017, sostienen que el silenciamiento de AQP5 restringió la migración e invasión de las células de cáncer colorrectal y regulaba la expresión de las células relacionadas con el proceso de transición epitelial-mesenquimal (EMT) en ellas al inhibir la vía de Wnt / b-catenina.

El proceso de EMT es la base molecular de la metástasis del cáncer. La EMT es un proceso que transforma las células epiteliales en células mesenquimales, y por lo tanto, las células epiteliales pierden su polaridad y adhesión celular y adquieren capacidades migratorias e invasivas [36]. La EMT generalmente se considera un factor principal en la fase temprana de las metástasis de cáncer [37]. La aparición de metástasis a distancia es principalmente un evento letal durante la progresión del tumor, y la metástasis hepática es más frecuente en CCR [38][39]. Los mecanismos moleculares detrás de la EMT del CCR han recibido mucha atención en los estudios recientes con la esperanza de identificar nuevos objetivos terapéuticos para una terapia eficaz contra el cáncer.

Por lo tanto, teniendo en cuenta que AQP5 está sobre-expresado en las células de CCR, y su sobreexpresión se correlaciona con un mal pronóstico de los pacientes con CCR, se demostró que la eliminación de AQP5 inhibió la migración y la invasión de las células de CCR, e indujo la alteración de la expresión de genes relacionados con EMT mediante la regulación negativa de la vía Wnt / B catenina.

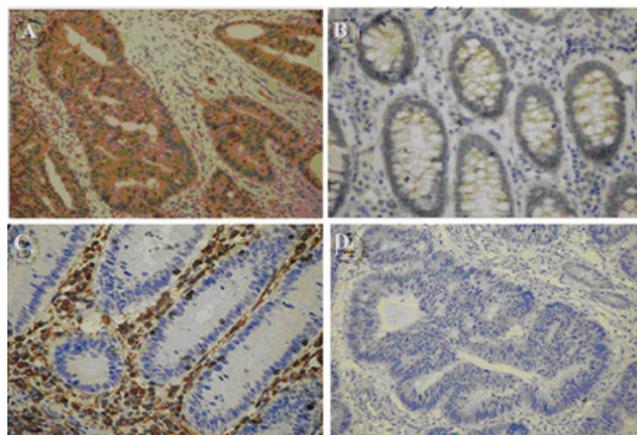


Figura 8: Imágenes representativas de tinción inmunohistoquímica para detectar la expresión de AQP5 y AQP8 en muestras clínicas CCR. **A** Inmunorreactividad AQP5 en el citoplasma y la membrana plasmática de las células de carcinoma colorrectal (marrón oscuro). **B** La inmunoreactividad de AQP5 no se detecta en el epitelio de colon normal adyacente. **C** La inmunoreactividad de AQP8 se detectó en el citoplasma y la membrana plasmática del epitelio de colon normal adyacente (marrón oscuro). **D** La inmunoreactividad de AQP8 no se detecta en los tejidos de carcinoma colorrectal. (Aumento: 400X). (Wei Wang et al., 2012[33]).

5.5.3. CDC25A y CDC25B

Los genes CDC25 son fosfatasa activadoras del ciclo celular que regulan positivamente la actividad de la quinasa dependiente de ciclina. En particular, CDC25A controla la progresión a través de la fase S, mientras que CDC25B y CDC25C participan en el control de la transición de G2 a mitosis [40]. CDC25A gen pertenece al cromosoma 3 humano y CDC25B al cromosoma

20 humano. Teniendo en cuenta que la desregulación de la progresión del ciclo celular es una alteración evidente en los tumores malignos humanos, CDC25A y CDC25B son oncogenes que se sobreexpresan en una variedad de neoplasias malignas humanas. La sobreexpresión de CDC25A y CDC25B se ha demostrado en el linfoma no Hodgkin, los carcinomas humanos de mama y pulmón y los tumores de cabeza y cuello [41][42][43][44].

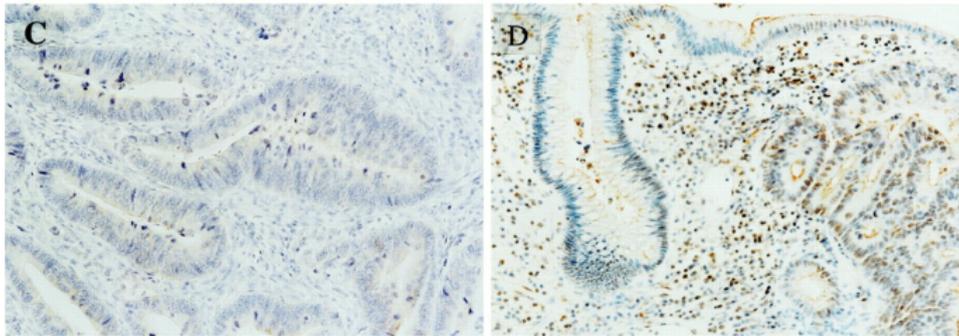


Figura 9: Inmunotinción con anticuerpo antihumano CDC25A. La proteína CDC25A se localizó principalmente en el núcleo tanto en tejido normal (C) como en tejido de carcinoma (D). Se ve una sobre-expresión de la proteína en tejido de carcinoma por sobre el epitelio normal. (Aumento de C: 25X)(Aumento de D: 50X) (Ichiro Takemasa et al., 2000[46]).

Por otro lado, el grupo de Rizeng Li sostiene que la sobre-expresión de CDC25B se observa en tejidos de CCR. Este resultado coincide con el análisis bioinformático realizado por el mismo grupo con la base de datos de TCGA [47]. A su vez, sostienen que la restauración de los niveles de expresión de CDC25B rescató significativamente la proliferación, migración e invasión de células CCR. La sobreexpresión de CDC25B hace que las células de la fase S y la fase G2 entren en mitosis, ya sea que la replicación del ADN esté completa o no. Por lo tanto, los altos niveles de esta proteína se asocian a la proliferación celular [48][49]. Diversos estudios asumen el rol oncogénico de CDC25B y aportan información de sobre-expresión en CCR. Ichiro Takemasa et al., indicaron que la alta expresión era un predictor significativo de mal pronóstico de CCR, y que su alta expresión se observaba en 43 % de casos colorrectales. A su vez, sostienen que hay una correlación significativa entre la alta expresión de CDC25B y el tamaño del tumor o metástasis a distancia. No hubo correlación entre la alta expresión de CDC25B y la edad, el género, el sitio del tumor, la metástasis de los ganglios linfáticos, la profundidad de la invasión y el grado de diferenciación [46]. En particular, la figura 10 representa la expresión de CDC25B, imagen adaptada de (Ichiro Takemasa et al., 2000)[46]. La figura A en la mucosa colónica normal, se observó una expresión débil de CDC25B desde la parte inferior hasta la parte superior del epitelio normal, mientras que la figura B, es un carcinoma colorrectal representativo que expresó un alto nivel de CDC25B.

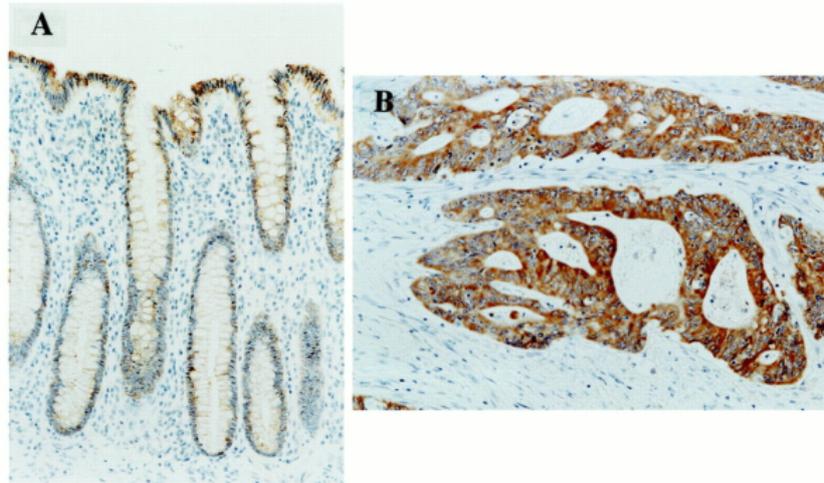


Figura 10: Inmunotinción con anticuerpo antihumano CDC25B. **A:** en mucosa colónica normal y **B:** tejidos de carcinoma colorrectal. A, en la mucosa colónica normal, se observó una expresión débil de CDC25B desde la parte inferior hasta la parte superior del epitelio normal. B, un carcinoma colorrectal representativo que expresó un alto nivel de CDC25B. La intensidad de la tinción se consideró fuerte (intensidad = 3), y el porcentaje de células CDC25B positivas fue del 100%. (Aumento de A: 25X)(Aumento de B: 50X). *Ichiro Takemasa et al., 2000[46]*.

5.5.4. FSCN1

Las fascinas son un grupo de proteínas que participan en el mantenimiento de una estructura adecuada del citoesqueleto celular. En particular, Fascin1 (FSCN1) es una proteína de agrupamiento de actina presente en mesenquimales, dendríticas, neuronas, músculo liso, fibroblastos y células endoteliales y se caracteriza por su ausencia en la mayoría de las células epiteliales. Los resultados de sobreexpresión en células de CCR se observan en Yosuke Hashimoto et al., 2006 y 2009 [50][51]. Dado que el crecimiento tumoral y las metástasis no son posibles sin un proceso altamente coordinado que involucra factores intracelulares y extracelulares, la migración de células de adenocarcinoma es posible por los cambios en la organización del citoesqueleto de actina, así como la matriz celular e interacciones adhesivas célula-célula. Además, los resultados en Jawhari et al., 2003, también reflejan que FSCN1 no es detectable en el epitelio colónico normal, mientras que en el adenocarcinoma colorrectal está sobre-expresado. La figura 11 es un ejemplo del estudio de Barbara M. Piskor et al, que demuestra estos resultados[52]. Además, los estudios realizados por Barbara M. Piskor, Anna Pryczynicz et al, 2018, así como Aida U. Jawhari, Andrea Buda et al, 2003, indican que el aumento de la expresión de FSCN1 fue más común en el tipo clínico desfavorable de adenocarcinoma sin componente de la mucosa que sugiere un papel potencial de esta proteína en la histogénesis del cáncer de colon [52][53].

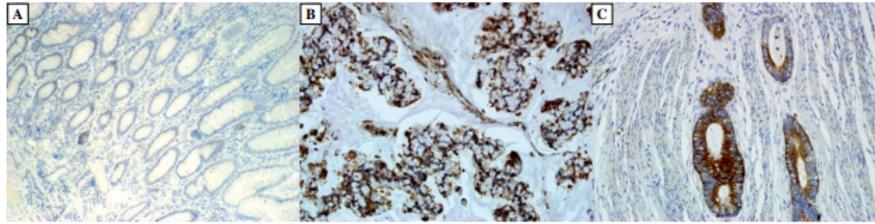


Figura 11: Expresión de la proteína FSCN1 en CCR. **A** Falta de inmunorreactividad de FSCN1 en la mucosa normal del colon. **B** Expresión positiva de FSCN1 en las células del CCR. **C** Inmunorreactividad con FSCN1 en glándulas similares a criptas de CCR. Aumento 200X. *Barbara M. Piskor et al., 2018 [52]*.

Se observó una expresión fuerte de FSCN1 en un subconjunto de adenocarcinomas colorrectales avanzados que se correlacionó con una supervivencia más corta en pacientes en estadio III y IV, con una expresión a menudo focal. Por lo tanto, la evidencia de numerosos estudios sugieren que FSCN1 puede tener valor-pronóstico como un biomarcador temprano para adenocarcinomas colorrectales más agresivos [51].

5.5.5. GLI1

GLI es miembro de la familia Kruppel, caracterizada por un dominio de proteína de dedo de zinc conservado. El gen GLI1 se encuentra en el cromosoma 12 de los humanos. El gen codificado media la señalización de Sonic hedgehog (Shh) durante el desarrollo de los mamíferos. Se sabe que la vía Shh desempeña funciones esenciales en múltiples aspectos del desarrollo embrionario, y su desregulación se ha implicado en la patogénesis de una variedad de tumores humanos [54]. Dado que juega un papel en la proliferación y diferenciación celular a través de su papel en la señalización de Shh, la evidencia acumulada ha demostrado que Gli1 está involucrado en la progresión tumoral [55][56]. La expresión de dicha proteína no está asociada a factores biológicos como la edad de los pacientes, el sexo, el tamaño del tumor y el grado tumoral. Por el contrario, la expresión de Gli1 aumenta en los tejidos de cáncer de colon en comparación con los tejidos normales, especialmente en pacientes con metástasis en los ganglios linfáticos y estadios T superiores [57]. Por otro lado, la señalización de Wnt-b-catenina desempeña un papel fundamental en la homeostasis del epitelio intestinal al promover la renovación de las células madre. En el intestino delgado, las células de Paneth epiteliales secretan ligandos Wnt y, por lo tanto, adoptan la función del nicho de células madre para mantener la homeostasis epitelial. Si bien no está claro cuáles células comprenden el nicho de células madre en el colon, se sabe que las células mesenquimales subepiteliales que expresan Gli1 forman este nicho esencial. El bloqueo de la secreción de Wnt de las células que expresan Gli1 previene la renovación de las células madre del colon en ratones: las células madre se pierden y, como consecuencia, la integridad del epitelio del colon se corrompe y conduce a la muerte [58]. El mismo grupo de estudio determinó que Gli1 se correlaciona con el cambio de tipo EMT de células epiteliales cancerosas en el adenocarcinoma colorrectal, al igual que AQP5, y que GLI1 se asocia con la proliferación de células cancerosas en el adenocarcinoma colorrectal. GLI1 promueve la potencia del cáncer a través de la vía de señalización intracelular PI3K / Akt / NFkB en el adenocarcinoma colorrectal. La figura 12 refleja la sobre-expresión de la proteína. La tinción de Gli1 fue detectable

tanto en el citoplasma como en el núcleo de las células epiteliales de colon benignas y malignas. En el núcleo de los tejidos normales del colon (Figura 12A), la expresión de la proteína Gli1 fue menos abundante que en el núcleo de las células de cáncer de colon. En las células del cáncer de colon, la expresión de Gli1 estaba presente en diferentes intensidades y diferentes distribuciones celulares. A su vez, se observó que el alto nivel de expresión de Gli1 estaba relacionado al alto estadio de la enfermedad [55].

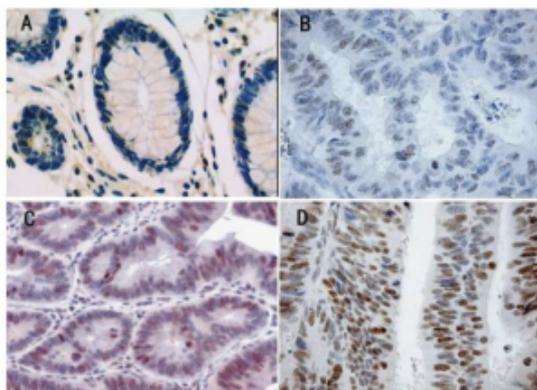


Figura 12: Análisis de expresión inmunohistoquímica de la proteína Gli1 en tejido de colon normal y tumores de colon utilizando un Microarray de tejidos (Aumento: 200 X). **A:** tejido de colon normal, Gli1 negativo. **B:** tejido de cáncer de colon, Gli1 positivo +, la tinción es débil. **C:** tejido de cáncer de colon, Gli1 positivo ++, la tinción es moderada. **D:** Tejido de cáncer de colon, Gli1 positivo +++, la tinción es fuerte. *Yin-Lu Ding et al., 2012 [55]*.

5.5.6. NQO1

Este gen pertenece al cromosoma 16 humano y codifica NAD (P) H: quinona oxidorreductasa (NQO1). Es una enzima antioxidante que protege a las células contra el estrés oxidativo y el ciclo redox. NQO1 es una flavoproteína que cataliza la desintoxicación y la reducción de quinonas en dos electrones, así como sus subproductos [59]. Algunos compuestos como xenobióticos, oxidantes y antioxidantes, la luz ultravioleta y las radiaciones ionizantes inducen la expresión del gen NQO1. El elemento de respuesta antioxidante (ARE) se activa por factores nucleares relacionados con E2, como el factor NF-E2 (Nrf2). Es un TF esencial en el manejo de la desintoxicación y la protección antioxidante, así como en la regulación general de la respuesta protectora. Nrf2 tiene la capacidad de gestionar muchos aspectos diferentes de la protección celular [60]. Los estudios realizados por Asma Chinigarzadeh, Razauden Zulkifli et al, 2012, indican que NQO1 también es un biomarcador potencial para el CCR. NQO1 está sobreexpresado en numerosos tumores sólidos humanos y altos niveles de NQO1 se correlacionan con un peor pronóstico del paciente. Altos niveles de NQO1 en los cánceres implica que las células cancerosas pueden tratar el estrés oxidativo elevado, y esto se confirma a partir de estudios de Nrf2, el TF que principalmente controla la expresión de NQO1. La hiperactivación de Nrf2 ayuda a las células transformadas o malignas a evadir el estrés oxidativo extremo a través de la inducción de genes citoprotectores, por ejemplo, la sobreexpresión de NQO1. Por lo tanto, la inhibición de NQO1 evitaría el crecimiento de células cancerosas y la transformación maligna, ya que dicha inhibición podría dar lugar a niveles tóxicos de estrés oxidativo[61].

5.5.7. VWA2

Este gen pertenece al cromosoma 10 humano y se conoce como colon cáncer secreted protein-2 (CCSP-2). La proteína codificada se localiza en la matriz extracelular y puede servir como un componente estructural en las membranas basales o en estructuras de anclaje en andamios de colágeno VII o fibrilina. Los estudios realizados por Beatriz González et al., 2018 y Baozhong Xin et al., 2005 confirman que VWA2 es un excelente biomarcador para el CCR, ya que su sobreexpresión indica especificidad de células cancerosas y de tejido. La proteína de secreción se encuentra ausente en tejido sano, tanto de colon como de otros, pero se encuentra sobreexpresada hasta 78 veces en CCR en estadios II a IV [62][63]. A su vez, el trabajo realizado por Kim et al., 2017, demostró la expresión de CCSP-2 en el 87.0% de las muestras de cáncer colorrectal y el 89.5% de las muestras de adenoma de colon, lo cual corrobora los estudios anteriores [64]. La figura 13 muestra los resultados de sobre-expresión de este estudio.

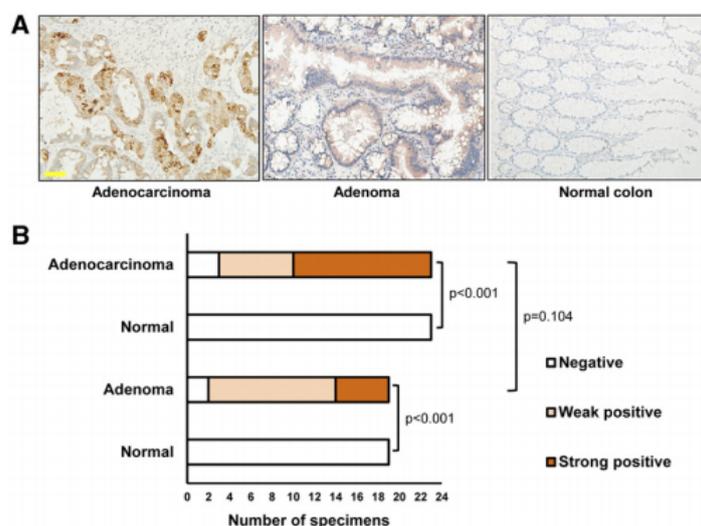


Figura 13: Se observó una expresión de CCSP-2 significativamente más fuerte en CCR y adenomas humanos en comparación con tejidos de colon normales emparejados. *Jaeil Kim et al., 2017 [64]*.

5.5.8. Promotores híbridos

Teniendo en cuenta todo lo mencionado, se afirma que la terapia génica del cáncer empleando virus oncolíticos deberían utilizar un vector viral capaz de infectar tumores con alta eficiencia, induciendo una expresión específica y de alto nivel de transgén en el tumor y destruyendo selectivamente las células tumorales[65]. Una posible solución para mejorar las terapias génicas es mejorar la eficiencia y la especificidad de los vectores en una de tres formas: con un mejor reconocimiento celular, con una entrada eficiente a la célula deseada, o con una expresión específica de tejido a niveles suficientes. Debido a que la expresión de las proteínas mencionadas en las secciones anteriores no son homogéneas en todos los tipos de CCR, y que un AOL construido por el PET de una única proteína limitaría su eficacia terapéutica a tumores que únicamente sobreexpresen dicha proteína, esto podría causar la proliferación de subclones que no la expresan.

Se propone la terapia con promotores híbridos. Los promotores híbridos, también conocidos como **promotores quimera**, son aquellos constituidos por elementos de distinto origen que buscan imitar las propiedades ideales de los vectores necesitados para la terapia génica en función del tipo de cáncer a tratar. Es decir, se busca armar los vectores con combinaciones de PET que logren diferenciar las células malignas de las normales según la expresión diferencial de genes de dichas células [66]. Teniendo en cuenta que es poco probable que los promotores mono-sensibles tengan un perfil de activación que refleje el curso de una enfermedad completa, se busca a través de los promotores quimera reflejar la activación de genes en distintos tipos celulares que componen la enfermedad.

Hay diversos estudios que utilizan promotores híbridos para lograr una actividad transcripcional mayor en un tejido determinado. Este es el caso del grupo de investigadores L Wu et al., quienes combinaron el promotor específico de la próstata (PSA) con elementos de respuesta a andrógenos (AR) activando la transcripción en presencia de estas hormonas [67]. A su vez, Katarzyna Piekarowicz et al., trabajaron en un nuevo promotor modular específico para músculos, el cual demostró su alta actividad en las células musculares y su baja actividad en otros tipos de células [66]. Por otro lado, O Greco et al., han logrado nuevos promotores de genes quiméricos sensibles a la hipoxia y la radiación ionizante [68]. Los investigadores Min Guan et al., presentan resultados de un nuevo vector híbrido de adenovirus-alfavirus, con mayor eficacia y seguridad en el tratamiento del cáncer de hígado experimental [65]. A su vez, diversos estudios han mostrado que el agregado de enhancers o activadores río arriba de un PET es útil para mejorar la expresión del transgen requerido para la terapia génica del tejido específico target [69].

Este nuevo tipo de vectores híbridos puede proporcionar una herramienta potente y segura para la terapia génica contra el cáncer. En esta tesis se busca identificar secciones de los promotores detectados por el laboratorio para lograr la mejor combinación de estos PET para direccionar transcripcionalmente al adenovirus oncolítico en CCR.

5.5.9. Promotores sintéticos

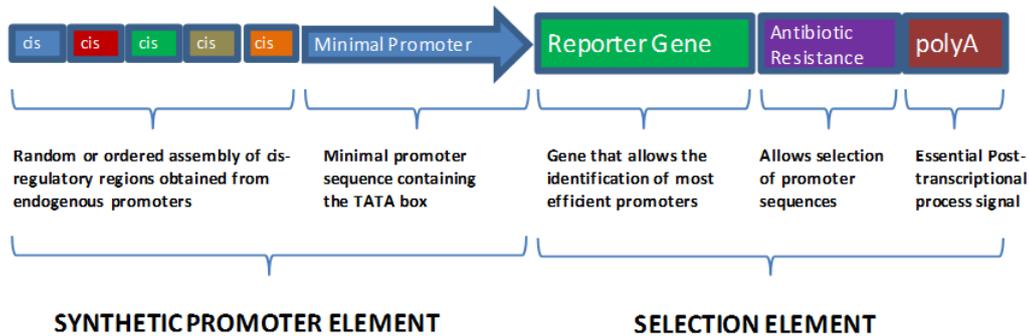
Los promotores sintéticos son secuencias de ADN que no existen en la naturaleza y que están diseñados para regular la actividad de los genes, controlando la capacidad de un gen para producir su propia proteína codificada de forma única. Cada gen requiere un promotor para decodificarlo en una proteína. Cuanto más fuerte es el promotor, más rápido se puede decodificar el gen y, como consecuencia, se puede producir más proteína. Teniendo en cuenta que el genoma está compuesto por diferentes promotores específicos para cada gen que han evolucionado para controlar la expresión génica de una manera particular, los promotores sintéticos comprenden fragmentos de dichos promotores naturales para formar nuevos tramos de secuencia de ADN que no existen naturalmente [70]. La generación de vectores de terapia génica sensibles a la enfermedad requiere el conocimiento del perfil de activación de los TFs durante la enfermedad activa, con el fin de ensamblar sitios de unión para estos TFs en promotores sintéticos, que pueden activarse de manera apropiada mediante el proceso de la enfermedad. Si bien los promotores sintéticos se encuentran en pleno desarrollo, diversos estudios buscan esclarecer su armado y funcionamiento. En particular, el grupo de Mohamed et al., se ha encargado de trabajar con el armado de promotores sintéticos para comprender distintos resultados. Dicho grupo sostiene que para armar los promotores sintéticos, sería lógico utilizar promotores

activados por varios TFs relevantes para el proceso de la enfermedad en estudio [71]. A su vez, los investigadores sostienen varias conclusiones en el armado de los promotores relacionados a:

- Actividad de promotores monosensibles en células expuestas a estimulación múltiple. Los promotores monosensibles deben diseñarse cuidadosamente para que retengan la actividad completa incluso cuando se activan otras vías transcripcionales o, como alternativa, se debe incorporar la capacidad de respuesta múltiple en los promotores para que otros factores de transcripción relevantes contribuyan a su activación.
- Disposición espacial óptima de 15 pb entre TFBS y 55 pb de la caja TATA el último TF para una baja actividad promotora basal e inducción máxima.
- La función de agrupaciones de TFBS espaciadas óptimamente y el efecto de la estimulación múltiple. Las repeticiones de los mismos TFBS en disposiciones cercanas promueve el reclutamiento del TF afín y, por lo tanto, aumenta la actividad de transcripción del promotor.

Por lo tanto, primero se debe conocer los TFs asociados a la enfermedad de cáncer colorrectal y su función en la misma, para luego decidir si los mismos se deben incorporar en el armado del promotor. A su vez, se ha estudiado que la incorporación de factores tales como SP1 es de vital importancia ya que la adición de sitios Sp1, sirven para la protección contra el silenciamiento del promotor a través de la metilación, así como también para impulsar la expresión génica [72]. A su vez, es muy importante determinar el espaciado de cada elemento del promotor de tal manera que los elementos reguladores aparezcan en el mismo lado de la hélice de ADN cuando se vuelven a ensamblar, y por lo tanto, deben emplearse elementos mínimos relevantes para que se produzcan así promotores capaces de expresarse eficientemente solo en el tejido de interés. Las secuencias reguladoras de cis derivadas de elementos de promotores naturales se usan para construir estos promotores sintéticos usando un enfoque de bloques de construcción, que puede llevarse a cabo mediante diseño racional o mediante ligadura aleatoria, tal como se muestra en la figura 14. En particular, el ensamblaje aleatorio de elementos reguladores cis ha mostrado un éxito particular como un medio para desarrollar promotores sintéticos [72]. El resultado es una secuencia de ADN compuesta de varios elementos reguladores cis distintos en una orientación completamente nueva que puede actuar como un potenciador del promotor; típicamente para iniciar la transcripción mediada por ARN polimerasa II.

A. Typical Mammalian Synthetic Promoter Layout



B. Typical Prokaryotic Synthetic Promoter



Figura 14: Diseños típicos de promotores sintéticos para procariontas y eucariotas [72].

A su vez, un importante avance realizado por (Mogno et al., 2010) indica que la caja TATA es un componente modular, ya que su fuerza de unión al complejo de ARN polimerasa II y la fuerza de transcripción resultante que media es independiente de los potenciadores del elemento regulador cis aguas arriba. También descubrieron que la caja TATA no agrega ruido a la transcripción, es decir, actúa como un amplificador simple sin alterar la especificidad de la expresión génica dictada por los elementos potenciadores aguas arriba. Por lo tanto, implica que cualquier combinación de potenciadores reguladores cis podría acoplarse a una caja TATA y serían los potenciadores los que mediarían en la especificidad sin ninguna interferencia de la caja TATA. Las implicaciones de este estudio sugieren que debería ser posible construir cualquier tipo de promotor sintético diseñado específicamente para mostrar un patrón altamente restrictivo de regulación génica, y por lo tanto, la construcción de promotores sintéticos es posible debido a la naturaleza modular de las regiones reguladoras de genes naturales [73].

Por otro lado, el grupo de investigación G M Edelman et al., mostraron que los promotores sintéticos mas activos en la línea celular de neuroblastoma estaban asociadas con un mayor número de factores conocidos, entre los cuales se encontraban: AP2, CEBP, GRE, EBOX, ETS, CREB, AP-1 y Sp1 / MAZ [74].

Una aplicación terapéutica interesante para la tecnología del promotor sintético que se ha descrito es la generación de una clase de virus competentes en replicación que permiten la destrucción específica de células tumorales al replicarse específicamente en células cancerosas. Uno de los precedentes más importantes que hay en esta rama es el estudio realizado por (Han et al., 2010), quienes desarrollo un promotor sintético específico del hígado que podría usarse para curar esencialmente la diabetes en un modelo de ratón transgénico. En este estudio se construyó un promotor sintético activo en células hepáticas en respuesta a la insulina. Los autores diseñaron promotores de 3, 6 y 9 elementos basados en combinaciones aleatorias de elementos cis HNF-1, E / EBP y GIRE. En los promotores de 3 elementos, se probaron las

27 combinaciones de las tres y se usaron los promotores de mayor actividad para generar el promotor de 6 elementos, y así sucesivamente [75]. Para que la terapia génica sea un éxito en la clínica, es imperativo desarrollar promotores que sean altamente específicos y que muestren un perfil de expresión restrictivo y predecible [72].

En esta tesis se busca diseñar promotores sintéticos que reemplacen el promotor natural de E1A y por lo tanto, se dirige así la actividad del adenovirus, logrando que el mismo se traduzca y replique eficientemente solo en las células cancerosas del tejido colorrectal. La eficiencia del promotor se basa en los TFs con los cuales se arma el promotor y la distancia de los mismos.

6. Bioinformática

La bioinformática es un campo interdisciplinario que involucra principalmente la biología molecular y la genética, la informática, las matemáticas y la estadística. La gran demanda de análisis e interpretación de estos datos está siendo gestionada por la ciencia en evolución de la bioinformática, y por lo tanto, se busca abordar a los problemas biológicos a gran escala con uso intensivo de datos desde un punto de vista computacional. Siendo un campo en pleno desarrollo, las técnicas computacionales son mayormente utilizadas con el fin de diseñar y avanzar en el campo de la medicina de precisión [76]. En dicha área, se busca definir el tratamiento, el tipo y el dosaje de medicación para el paciente en función de la información genética particular del mismo. Además del análisis de los datos de la secuencia del genoma, la bioinformática se está utilizando ahora para una amplia gama de otras tareas importantes, incluido el análisis de la variación y expresión génica, el análisis y la predicción de la estructura y función de genes y proteínas, la predicción y detección de redes de regulación génica, la simulación de entornos para el modelado de células completas, el modelado complejo de la dinámica y las redes reguladoras de genes, y la presentación y análisis de vías moleculares para comprender las interacciones gen-enfermedad. Aunque a menor escala, las tareas bioinformáticas más simples y valiosas para el investigador clínico pueden variar desde el diseño de cebadores (secuencias cortas de oligonucleótidos necesarias para la amplificación de ADN en experimentos de reacción en cadena de la polimerasa) hasta la predicción de la función de los productos génicos. En todos los casos, una vez obtenidos los datos necesarios se hacen inferencias a partir de los mismos [77].

El **proyecto genoma humano** es uno de los trabajos más importantes alcanzados en esta área. El mismo implicó un programa de investigación colaborativo e internacional cuya meta era la del mapeo y entendimiento completo de todos los genes de los seres humanos. Todos nuestros genes juntos se conocen como *genoma*. Una vez finalizado el proyecto en el 2003, se dio a conocer de la existencia de aproximadamente 25000 genes en el genoma humano, brindando así un recurso de información detallada acerca de la estructura, la secuencia de bases de nucleótidos, la localización, la organización y la función del conjunto completo de genes humanos [78]. Los tres grandes cambios que conllevó este proyecto son: la producción de datos a gran escala, el desarrollo de recursos computacionales potentes y robustos y la sociología de la investigación biológica, ya que la magnitud y complejidad de los proyectos genómicos requieren de la colaboración en grandes redes científicas, con equipos multidisciplinares complementarios. Se considera que el siglo XXI, uno de los mayores paradigmas tras el legado del proyecto genoma humano es el de la reconstrucción de los organismos, el de un enfoque sintético, inter-

disciplinario, donde grandes grupos de investigación formados por genéticos, bioinformáticos, bioquímicos, físicos, matemáticos, médicos, informáticos, trabajan conjuntamente para explicar las propiedades emergentes de los sistemas biológicos [79].

Dos grandes áreas de la bioinformática son **minería de datos** y **minería de texto**.

6.1. Minería de datos

Las bases de datos biológicas son el corazón de la bioinformática y representan un conjunto organizado de una gran variedad de datos biológicos de investigaciones anteriores realizadas en laboratorios (incluidos estudios *in vitro* e *in vivo*), desde análisis bioinformáticos (*in silico*) y artículos científicos. El análisis de grandes volúmenes de datos biológicos a través de sistemas de bases de datos tradicionales es a menudo problemático y desafiante. La minería de datos es el proceso de estructurar el conocimiento y los patrones de los grandes volúmenes de estos datos. Teniendo en cuenta el crecimiento de los datos biomédicos, la necesidad de clasificación de datos útiles al proyecto de cada investigador se vuelve una etapa elemental en el flujo de trabajo. A través de algoritmos innovadores de clustering, regresión, clasificación, y redes neuronales se busca mejorar las técnicas de minería de datos innovadoras e inteligentes para un análisis de biodatos efectivo y escalable [80]. Bases de datos actuales como RefSeq, The Human Protein Atlas (THPA), EMBL, Protein Data Bank (PDB), Ensembl, Gepia, entre otros, son grandes fuentes de información, y la complejidad acarreada por los grandes volúmenes de datos es lo que se intenta abordar con la minería o extracción de datos [81].

6.2. Minería de texto

Teniendo en cuenta que los científicos informan sus resultados en lenguaje natural, hay un desafío tanto en el análisis computacional de texto como en la creación de bases de datos estructuradas de información reportada usando lenguaje natural. Por lo tanto, el objetivo de la minería de texto es proporcionar una revisión útil de los datos para el investigador. Este proceso implica la aplicación de técnicas de procesamiento del lenguaje natural para la obtención de información biológica relevante a partir de las plataformas tales como *PubMed*, que contienen los respectivos datos biomédicos [82][83]. Un ejemplo de minería de texto se observa con el tesoro Medical Subject Headings (MeSH), un vocabulario controlado y jerárquicamente organizado producido por la Biblioteca Nacional de Medicina. Se utiliza para indexar, catalogar y buscar información biomédica y relacionada con la salud. MeSH incluye los encabezados de temas que aparecen en MEDLINE / PubMed, el Catálogo NLM y otras bases de datos NLM [81].

6.3. Bases de datos

A continuación, se desarrollan las bases de datos de TRANSFAC y JASPAR en mayor detalle ya que son las que fueron empleadas para el análisis de este proyecto.

6.3.1. TRANSFAC

TRANSFAC® es la base de datos de TFs eucariotas, sus sitios de unión genómica y sus perfiles de unión al ADN. Se remonta a una compilación muy temprana, se ha mantenido y conservado cuidadosamente desde entonces y se convirtió en el estándar de oro en el campo, que se puede utilizar al aplicar la plataforma geneXplain. En particular, su biblioteca de matrices de peso posicional es una colección única de modelos de unión al ADN, adecuada para un análisis exhaustivo de secuencias genómicas para posibles TFBS. TRANSFAC® se puede utilizar como enciclopedia de regulación transcripcional, o como una herramienta para identificar posibles TFBS. La plataforma online de MATCH desarrollada en la sección de *Materiales* permite hacer esto último.

La estructura de TRANSFAC consta de un núcleo que comprende el contenido de dos dominios: uno documenta los sitios de unión de TFs, generalmente en promotores o potenciadores. El otro describe las proteínas de unión. La figura 15 representa dicho esquema. Los sitios de unión que se refieren al mismo TF se fusionan en una matriz de peso posicional. Dicha matriz refleja la frecuencia con la que se encuentra cada nucleótido en cada posición de los sitios de unión de este TF y, por lo tanto, la preferencia de base en cada posición.

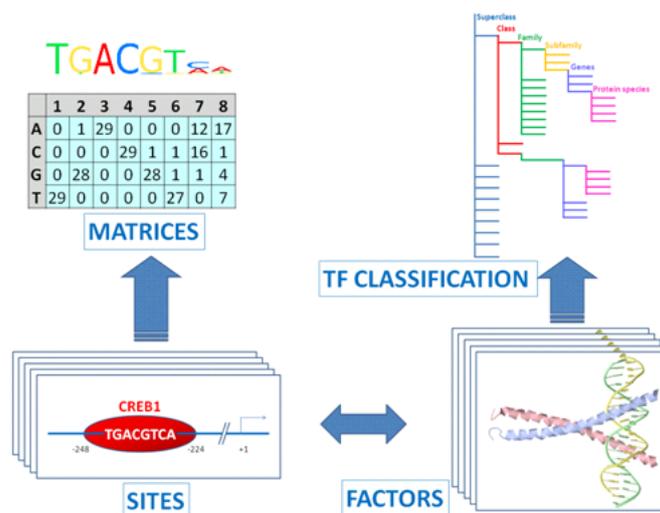


Figura 15: Estructura de la base de datos TRANSFAC

Las características principales de esta base de datos son:

- Los informes entrelazados que conectan TFs, sus sitios de unión caracterizados experimentalmente y genes regulados, así como informes de promotores con sitios de unión de TFs anotados mapeados y datos de alto rendimiento (ChIP-seq).
- Más de 23,000 informes de TFs (y 1,200 miRNA), un subconjunto de los cuales proporciona asignaciones funcionales GO, asociaciones de enfermedades y asignaciones de patrones de expresión.
- Más de 70,000 informes de sitio que contienen detalles de la literatura primaria para más de 300 especies, con un enfoque en humanos, ratones, ratas, levaduras y plantas.

- Más de 67,000 interacciones anotadas manualmente de factor de transcripción-sitio.
- Más de 7,000 matrices de peso posicional, para ser utilizadas por la plataforma online geneXplain MatchTM, FMatch, CMsearch que permiten predecir los sitios de unión de TF.
- Más de 360,000 informes de promotores para humanos y otros nueve organismos, incluidos sitios de inicio de transcripción, islas CpG, polimorfismos de un solo nucleótido (SNP) y varias otras anotaciones.
- Incluye herramientas para la predicción de sitios de unión a TFs, identificación de factores de novo, comparación de matrices e identificación de reguladores de miARN.

Uno de los grandes beneficios de la base de datos es que ofrece una plataforma online - geneXplain- con una versión gratuita de las matrices de peso. A su vez, existen plataformas online, al margen de la versión gratuita que ofrece TRANSFAC, que implementan sus propias matrices de peso posicional para predecir los sitios de unión del TF dentro de una secuencia de ADN, es decir, empleando sus propios algoritmos de búsqueda utilizando la base de datos TRANSFAC.

Sin embargo, se debe tener en cuenta que la versión completa de TRANSFAC (Professional) es una base de datos que se debe pagar. La versión utilizada en la plataforma online es una versión con contenido limitado. La misma base de datos compara la versión open source de la privada. Como bien aclara la plataforma online, el contenido de la versión profesional está más de diez años antes del lanzamiento público e incluye datos y capacidades de análisis que no están disponibles en la versión pública. La figura 16 es la información que se dispone de ambas versiones de la base de datos.

TRANSFAC®	Professional (version 2020.1)	Public
Data		
Factors	48,084	6,133
miRNAs	1,771	n/a
DNA sites	50,912	7,195
mRNA sites	67,823	n/a
Factor-DNA site links	68,928	Yes
miRNA-mRNA site links	74,676	n/a
Genes	102,900	2,397
ChIP-Seq fragments	103,548,181	n/a
Matrices	9,962	398
References	40,648	Yes
Promoter sequences	441,771	n/a

Figura 16: Información obtenida en Gene Regulation comparando TRANSFAC Professional y TRANSFAC public.

Aun así, es importante aclarar que a los fines del análisis del proyecto, los 6133 TFs de la base de datos pública son suficientes para hacer un análisis exhaustivo de las secuencias de los promotores. De esta manera, se permite realizar un análisis completo ya que se puede complementar y comparar los resultados con esta base de datos.

6.3.2. JASPAR

JASPAR (<http://jaspar.genereg.net>³) es una base de datos de acceso abierto de perfiles de unión a los TFs conservados y almacenados como matrices de frecuencia de posición (PFM) y modelos flexibles TF (TFFM) para TFs en múltiples especies en seis grupos taxonómicos. La base de datos JASPAR consta de subconjuntos más pequeños de perfiles conocidos como colecciones. Cada una de estas colecciones tiene objetivos diferentes. La colección principal se conoce como JASPAR CORE y es la colección que usan la mayoría de los científicos. La misma contiene un conjunto curado, no redundante, de perfiles de unión de TFs. Todos los perfiles se derivan de colecciones publicadas de sitios de unión de TFs definidos experimentalmente para eucariotas multicelulares. Uno de los objetivos centrales de JASPAR CORE es proporcionar un modelo único y *mejor* para cada TF. Esto significa que la base de datos no es redundante en el sentido de que no hay muchos modelos para el mismo factor (con algunas pocas excepciones motivadas por el reconocimiento de factores significativamente diferentes).

La principal diferencia con recursos similares (TRANSFAC - quien ofrece una versión pública con resultados parciales de la base de datos privada) consiste en el acceso a datos abiertos, la no redundancia y la calidad de JASPAR. Es importante aclarar que JASPAR CORE es lo que la mayoría de los científicos quieren decir cuando se refieren a JASPAR en manuscritos. Por conveniencia, JASPAR CORE está dividido por grupos más grandes de especies. Esta distinción se usa principalmente en la interfaz web y, opcionalmente, en la sección de descarga. Actualmente estos grupos taxonómicos más grandes son: vertebrados, planchas, insectos, nematodos, hongos, plantas y urocordados. En el directorio DESCARGAR, la mayoría de las colecciones de matrices tienen un subdirectorio SITE, que para cada modelo enumera todos los sitios utilizados para la construcción del modelo como un archivo fasta. Las alineaciones son implícitas: las subpartes de secuencias utilizadas están en mayúsculas. Se debe tener en cuenta que en la mayoría de los casos, esta es una interpretación: se utilizan buscadores de patrones para encontrar la alineación más probable, pero esto no siempre es lo más correcto. Esta es la razón principal por la que se pone a disposición las colecciones: de esta manera, les permiten a los usuarios crear sus propios modelos en función de los archivos sin formato.

Mas allá del formato de salida que se elija, cada matriz tiene asociada una detallada información haciendo clic en el ID de perfil y los logotipos de secuencia, mostrando información detallada sobre el modelo: tanto datos de anotación, como un logotipo de secuencia, una matriz de conteo y estadísticas de matcher (hits). Los logotipos de secuencia son representaciones gráficas del modelo de matriz, basadas en el contenido de la información. Un logotipo de secuencia es básicamente un diagrama de barras que muestra el contenido total de información en cada posición, donde la barra se reemplaza por letras apiladas (A, C, G, T), que se clasifican en función de su aparición. La figura 17 es un ejemplo de la matriz MA0048.2 asociada al TF NHLH1. Se presenta una descripción del factor, el logotipo de secuencia y la matriz de frecuencia, con opción de bajar la información en distintos formatos de salida.

³<http://jaspar.genereg.net>

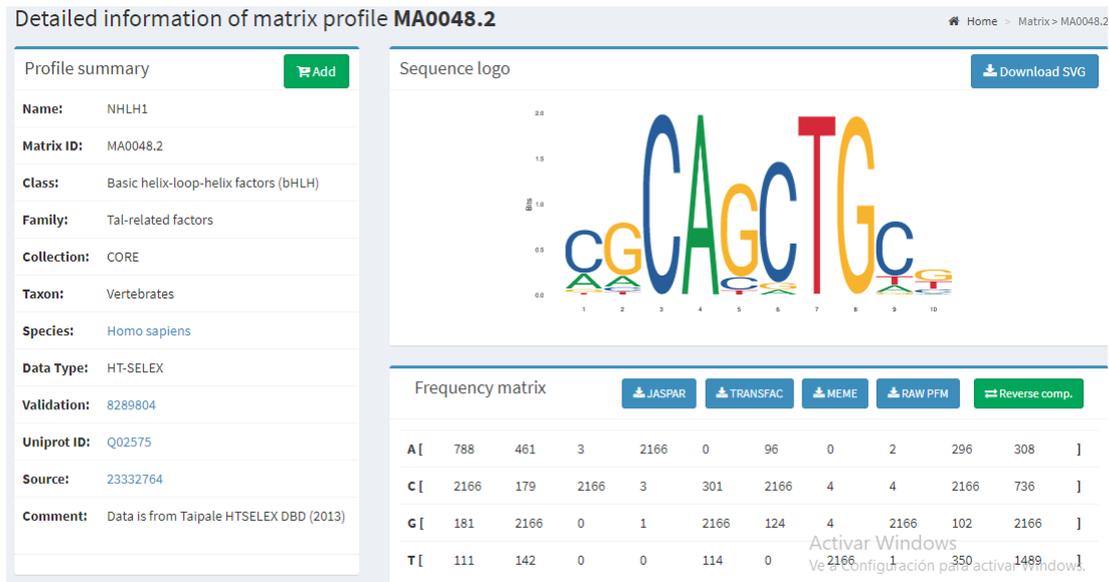


Figura 17: Formato de salida para una matriz asociada a un factor de transcripción.

Como bien muestra la figura 17, los PFM se pueden descargar en varios formatos, incluidos JASPAR, TRANSFAC y MEME. La diferencia entre ellos no es la información de la matriz, sino el diseño que tienen los resultados. En particular, el formato de factor MEME es un formato de texto simple para factores que es aceptado por los programas de MEME Suite que requieren el formato de factor MEME. Un archivo de texto en formato de factor mínimo MEME puede contener más de un factor, y también (opcionalmente) especifica el alfabeto del factor, las frecuencias de fondo de las letras en el alfabeto y la información del filamento (para factores de alfabetos complementarios como el ADN). Este es el formato que se elige para este proyecto ya que una de las plataformas online que se va a utilizar para analizar las secuencias de los promotores es un programa MEME Suite.

En este proyecto, se trabaja con la opción de *Download JASPAR data*, seleccionando la opción de *non-redundant*, y la opción de MEME, tal como indica la figura 18.

Taxonomic group	PFMs (non-redundant)	PFMs (redundant)
Vertebrates	<ul style="list-style-type: none"> Individual PFMs (zip) [JASPAR MEME TRANSFAC] Single batch file (txt) [JASPAR MEME TRANSFAC] 	<ul style="list-style-type: none"> Individual PFMs (zip) [JASPAR MEME TRANSFAC] Single batch file (txt) [JASPAR MEME TRANSFAC]

Figura 18: Selección de matrices de frecuencia de posición (PFM); grupo taxonómico: Vertebrados; *Single batch file (txt)* en formato MEME.

Toda esta información de la base de datos fue encontrada en la página oficial de JASPAR [84].

6.3.3. TRANSFAC vs JASPAR

La diferencia fundamental entre ambas es que JASPAR es de código abierto, mientras que TRANSFAC es un producto comercial de acceso gratuito de propiedad de BioBase (pertenece al grupo de Qiagen). Los TFs se pueden encontrar usando una variedad de métodos

que incluyen algoritmos que hacen el descubrimiento de dichos TFs a partir de secuencias que contienen sitios de unión y métodos *in vitro* como microarrays de unión a proteínas (PBM) y evolución sistemática de ligandos de alto rendimiento mediante enriquecimiento exponencial (HT-SELEX). Las mejoras en los procesos de secuenciación y las técnicas como la inmunoprecipitación de cromatina (ChIP) seguida de la secuenciación profunda (ChIP-seq) y la escisión de exonucleasa en ChIP-exo que miden la ocupación de TFs *in vivo*, han mejorado la resolución al nivel de un solo nucleótido. Todos estos métodos permiten proporcionar datos de alta resolución para el descubrimiento de factores. Teniendo esto en cuenta, las bases de datos mencionadas utilizan estas técnicas y obtienen información de manera independiente una de a otra. Por lo tanto, como la implementación de la información disponible es distinta, los resultados de las bases de datos no son necesariamente iguales.

JASPAR también afirma que tienen una calidad superior de curación y menos redundancia. Sin embargo, no se puede asegurar que tan verídico es esto, ya que TRANSFAC está considerada como una fuente altamente confiable y uno de sus puntos de venta para la versión profesional es que se paga por la calidad profesional de sus datos y herramientas. De todas maneras, es intrínsecamente especulativo ya que no hay forma de comparar los dos en cuanto a cuál es “mejor”, y si es así, ¿en qué sentido de la palabra?. Las medidas de calidad para medir esto son en gran medida subjetivas.

Es importante mencionar que la versión gratuita de TRANSFAC está desactualizada y es muy limitada. Sin embargo, es una fuente más de información que se debe analizar con ojo crítico y sirve para complementar los resultados obtenidos. Los papers analizados que tienen como objetivo identificar TFs, coincidente con el planteado en este trabajo, utilizan por lo general ambas bases de datos [85][86][87].

Tompa et al, en el 2005 probó algoritmos de descubrimiento de factores por su capacidad de predecir sitios de factores insertados usando medidas estadísticas para la sensibilidad del sitio y el coeficiente de correlación [88]. En este primer estudio exhaustivo, encontraron que un problema de evaluación de factores es complejo y admitieron que la inserción de factores aleatorios no captura la condición biológica de la unión de TF. Caleb Kipkurui Kibet y Philip Machanick muestran que ninguna base de datos única actualmente supera a las demás para todos los TF. Existe un acuerdo en el ranking de los mejores DB (ZLAB y HOCOMOCO) y peor desempeño (TF2DNA y SWISSREGULON), según se estudio realizado en 2016[89].

Lo que es importante mencionar es que en este proyecto no se busca identificar si existe una plataforma online mejor que la otra. Si se debe analizar los algoritmos de búsqueda y las distintas fuentes de información que han utilizado para justificar sus resultados. Sin embargo, se busca complementar los resultados con el fin de enriquecer el análisis realizado.

7. Hipótesis

Los diversos estudios ya publicados por el laboratorio de Osvaldo Pordhajcer del *Instituto Leloir* demuestran que es posible generar adenovirus oncolíticos capaces de eliminar células malignas con mayor eficiencia a partir de la regulación transcripcional de la replicación viral por medio de PETs. En particular, esto se corroboró con el AV22EL, regulado por el PET A33, y el AR2015, regulado por 2 PETs de genes que se expresan de manera heterogénea tanto en tumores

primarios como en metástasis de CCR (A33 y VWA2). Sin embargo, los adenocarcinomas de colon son altamente heterogéneos y por lo tanto, todos los tumores son distintos, y dentro de esta diferencia, no todos los tumores expresan A33 o VWA2. Por lo tanto, es esperable que aquellas células que no expresen alguna de las dos proteínas pueden escapar al tratamiento adenoviral y derivar en la recidiva de la enfermedad. Se puede mejorar el tratamiento génico mediante la propuesta de promotores híbridos y promotores sintéticos que logren una mejor respuesta en una mayor cantidad de células tumorales, y pueden ser identificados a través de la bioinformática, con herramientas de minería de datos y texto, haciendo inferencias a partir de los mismos.

8. Objetivo General

Este proyecto tiene por objetivo realizar un análisis bioinformático de los factores de transcripción de los 8 PETs detectados por el laboratorio, identificar aquellos que tienen una relevancia en la regulación de CCR para realizar en un futuro la construcción de nuevos AOL-RC. Por lo tanto, se centra en la regulación transcripcional del adenovirus. Por un lado, se estudian las secuencias de los PETs detectados en el laboratorio para combinar distintos PETs, identificando las fracciones de mayor interés para formar promotores híbridos. A su vez, se busca identificar factores de transcripción específicos de CCR para el armado de promotores sintéticos.

9. Objetivos Específicos

9.1. Promotores híbridos

1. Analizar los factores de transcripción de cada uno de los promotores y las distancias al inicio de la transcripción.
2. Analizar similitudes y diferencias entre los factores de transcripción de los 8 promotores.
3. Proponer secuencias de los promotores para combinarlos en promotores híbridos a partir de 1) y 2).

9.2. Promotores sintéticos

1. Proponer factores de transcripción relevantes para el armado de promotores sintéticos asociados a la enfermedad de cáncer colorrectal.
2. Proponer promotores sintéticos con combinaciones aleatorias de los factores de transcripción encontrados para dirigir la terapia de cáncer colorrectal.

Materiales y Métodos

10. MATERIALES

10.1. Plataformas online

10.1.1. ToppGene

ToppGene (<https://toppgene.cchmc.org/>⁴) es un portal integral gratuito y abierto a todos los usuarios, y no requiere un inicio de sesión para acceder. Esta plataforma online es única tanto para enriquecimiento funcional de la lista de genes, como para la priorización de genes candidatos mediante anotaciones funcionales o análisis de red de interacciones proteicas. La figura 19 [90] es una representación esquemática del flujo de trabajo y la metodología en las aplicaciones de ToppGene Suite. El funcionamiento de ToppGene asume que las características fenotípicas similares se asocian con genes que tienen funciones similares. El programa compara dos listas provistas por el usuario, una de **genes de entrenamiento** y una de **genes de prueba** usando anotaciones funcionales de todo el genoma. De esta manera, realiza un ranking **priorizando los genes candidatos** de la enfermedad basada en la anotación funcional, utilizando una medida de similitud entre dos genes basada en anotaciones semánticas. Los puntajes de similitud de las características individuales se combinan en un puntaje general mediante el metanálisis estadístico, teniendo un único valor-p ajustado para cada resultado [90]. Hay catorce puntajes de similitud descriptas en la figura 19 C y D.

⁴<https://toppgene.cchmc.org/>

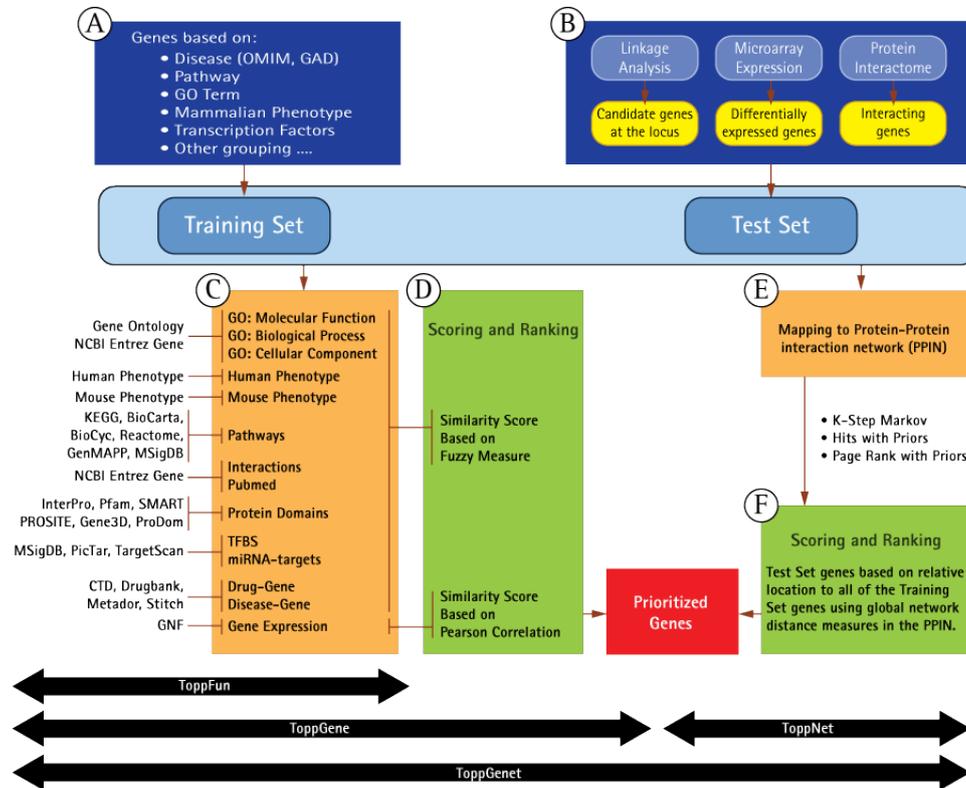


Figura 19: La **figura A** describe a los genes en el conjunto de capacitación que se seleccionan en función de sus atributos o anotaciones genéticas actuales (genes asociados con una enfermedad, fenotipo, vía o término GO). La **figura B** describe la fuente del gen de prueba, que pueden ser genes candidatos de estudios de análisis de ligamiento o genes expresados diferencialmente en una enfermedad o fenotipo particular. La **figura C** describe ToppFun: términos enriquecidos de las anotaciones genéticas y las características de la secuencia, a saber, GO: función molecular, GO: proceso biológico, fenotipo de ratón, vías, interacciones de proteínas, dominios de proteínas, sitios de unión del TF, genes objetivo de miRNA, enfermedad- asociaciones genéticas, interacciones farmacológicas-gen y expresión génica, compiladas a partir de varias fuentes de datos, y también utilizadas para desarrollar un perfil genético de conjunto de entrenamiento. La **figura C y D** describe la herramienta ToppGene: se genera un puntaje de similitud para cada anotación de cada gen de prueba comparándolo con los términos enriquecidos en el conjunto de genes de entrenamiento. La lista final de genes priorizados se calcula en función de los valores agregados de los catorce puntajes de similitud. La **figura E y F** describen la herramienta ToppNet: los genes del conjunto de entrenamiento y prueba se asignan a la red de interacción proteína-proteína, y realiza una puntuación y clasificación de los genes del conjunto de prueba según la ubicación relativa de todos los genes del conjunto de entrenamiento.

Finalmente, luego de haber realizado el análisis de comparación, se obtiene una tabla en formato csv. La misma contiene los genes de la lista de prueba ingresada por el usuario ordenados de menor a mayor en función del valor-p asignado según los puntajes mencionados calculados con la lista de entrenamiento. Por otro lado, se obtiene información de los genes de la lista de entrenamiento.

10.1.2. Gene Regulation - Match

La plataforma online llamada *Gene Regulation - Match* (<http://gene-regulation.com>⁵) utiliza la base de datos TRANSFAC, la cual brinda información de elementos reguladores de elementos reguladores de células eucariotas cis y trans, tal como se mencionó en la sección *Introducción*. Por lo tanto, proporciona un conjunto de datos que indican:

- Factores de transcripción eucariotas
- Sitios de unión probados experimentalmente: No solo proporciona los nombres de los factores de transcripción asociados a una secuencia, también brinda la posición en la cual se encuentra dicho factor
- Secuencias de unión de consenso: presenta la secuencia que representa al TF
- Genes regulados

La plataforma permite elegir distintos programas, entre los que se encuentran Match, F-Match, Patch, P-Match, MatrixCatch, AliBaba2, Composite Module Analyst (CMA), molwSearch, 3DTF, m2match, SignalScan, SbBlast, TfBlast, VCFmerge. En particular para este proyecto se eligió trabajar con Match. El mismo es un programa que utiliza la biblioteca de matrices de peso posicional de TRANSFAC® Public 6.0 para predecir TFBS en secuencias de ADN brindadas por el usuario. Las matrices pueden ser de alta calidad y de calidad media o baja, y la plataforma permite seleccionar la opción de trabajar únicamente con matrices de alta calidad. Para obtener los resultados, primero se debe crear un usuario de manera gratuita. Una vez que se tiene acceso, se debe ingresar en formato FASTA la secuencia que se desea analizar. A continuación, se debe seleccionar el grupo de matrices con los cuales el algoritmo de búsqueda va a trabajar. Las opciones son *all*, *vertebrates*, *fungi*. Luego, la opción de *high quality matrices only* permite aplicar el filtro mencionado previamente. A su vez, la opción de *perfiles* permite dirigir el algoritmo a un grupo de matrices determinados por *immune cell specific*, *cell cycle specific*, *liver specific*, *muscle specific*, *best selection*. A su vez, permite elegir la estrategia de filtrado para el corte del grupo matriz, ya sea minimizar los falsos positivos, minimizar los falsos negativos, minimizar la suma de ambas tasas de error o la opción de que el usuario ingrese valores de corte en función de los parámetros de probabilidad con que evalúan los resultados que explicaré a continuación.

Una vez elegidos dichos parámetros, los resultados se presentan en formato de tabla donde la primer columna es la matriz con la cual se identifica, la posición y la hebra (positiva o negativa), dos valores de probabilidad llamados *core match* y *matrix match*. Por un lado, *core match* es un valor entre 0 a 1 que indica como se adecua el factor de la base de datos a la

⁵<http://gene-regulation.com>

secuencia elegida. Por otro lado, *matrix match* es un valor entre 0 a 1 que indica como se dispone el factor a la position weight matrix. Por lo tanto, un resultado óptimo es que ambos valores sean 1. Luego, la columna de *sequence* presenta el extracto que se ha matcheado con el factor. Los nucleótidos que están en mayúscula son aquellos que coinciden 100 % con la matriz identificada, el resto de los nucleótidos se identifican con la secuencia en un cierto porcentaje en función de la matriz elegida. Todo esto se resume en los valores de *core match* y *matrix match* que la plataforma brinda para cada TF. Finalmente, la columna de *Nombre del Factor*, indica el nombre del TF. Por lo tanto, se obtiene un conjunto de factores con sus resultados de probabilidad correspondientes. La plataforma permite visualizar los resultados en un gráfico. En la plataforma hay una opción de *Goto Match Profiler* en la cual se encuentra la lista de matrices asociadas a los factores de transcripción con las cuales trabaja la plataforma. A su vez, cada una de ellas tiene asignado *high quality* o *low quality*. Las matrices de baja calidad son las que se filtran en caso de que el usuario lo desee [91].

10.1.3. Promo

PROMO (<http://algggen.lsi.upc.es/>⁶) es un laboratorio virtual para la identificación de supuestos TFBS en secuencias de ADN de una especie o grupos de especies de interés. Los TFBS definidos en la base de datos TRANSFAC se usan para construir matrices de peso de sitios de unión específicos para la predicción de TFBS. El usuario puede inspeccionar el resultado de la búsqueda a través de una interfaz gráfica y archivos de texto descargables. La plataforma online PROMO utiliza la versión de TRANSFAC 8.3. A su vez, la plataforma aclara que la misma es un proyecto de investigación colaborativo entre el Dr. M.M. Albà y D. Farré en el Grupo de Genómica Evolutiva (Unidad de Investigación en Informática Biomédica, Universitat Pompeu Fabra) y D. García, R. Escudero, O. Nuñez, J. Martínez y Dr. X. Messeguer de la Universitat Politècnica de Catalunya. Este sitio está parcialmente respaldado por el proyecto DGES PB98-0926 (AEDRI; Algoritmos y estructuras de datos para la recuperación de información). La variabilidad de la secuencia intrínseca de los sitios de unión a factores puede representarse como matrices de peso. Las mismas almacenan información sobre la frecuencia de los diferentes nucleótidos a lo largo del sitio, generalmente son más específicos (menos falsos positivos) y permiten calificar las coincidencias. PROMO construye matrices de peso que representan sitios de unión de factores específicos identificados experimentalmente en una especie particular o grupo de especies. La fuente de secuencias de sitio y TFs es TRANSFAC. Para buscar los posibles sitios en la secuencia de interés, la plataforma detalla los pasos a seguir.

1. *SelectSpecies*: selecciona la especie o grupo de especies de interés. También se puede seleccionar factores para restringir la predicción a un conjunto de TFs.
2. *SearchSites*: ingresar una secuencia de consulta para buscar posibles sitios de enlace.

Una vez que se ha seleccionado la especie, se construyen matrices de peso derivadas de al menos tres sitios de unión diferentes por TF (en la sección HELP PROMO se explica el procedimiento para construirlas). La construcción de la matriz se realiza sobre la marcha, lo que puede demorar algunos minutos. Sin embargo, en muchos casos las matrices ya se han calculado previamente (por ejemplo, para humanos o para todos los organismos, como es en nuestro caso con los

⁶<http://algggen.lsi.upc.es/>

promotores), por lo que no hay necesidad de esperar. El algoritmo busca la subsecuencia (núcleo) completamente conservada más larga para anclar la alineación de los sitios de unión. Se puede descartar un subconjunto de secuencias si esto conduce a un aumento en el número de sitios completamente conservados, siempre que al menos la mitad de las secuencias originales se mantengan en la alineación final. Cualquier secuencia terminal que contenga huecos se elimina. Las matrices de peso que contienen el número de ocurrencias de cada nucleótido en cada posición de secuencia se derivan de las alineaciones. Se pueden inspeccionar utilizando el módulo *ViewMatrices*, que también muestra los valores esperados en diferentes niveles de similitud y una lista de genes que contienen los sitios representados en la matriz [92].

10.1.4. Meme Suite- Fimo

Meme Suite - Motif-based sequence analysis tools (<http://meme-suite.org/index.html> ⁷) es una herramienta que permite descubrir nuevos factores en colecciones de secuencias de nucleótidos o proteínas no alineadas, y realizar una amplia variedad de otros análisis basados en factores [93]. MEME Suite admite análisis basados en factores de secuencias de ADN, ARN y proteínas. Proporciona algoritmos de descubrimiento de factores utilizando modelos probabilísticos y discretos, que tienen fortalezas complementarias. También permite descubrir factores con inserciones y eliminaciones arbitrarias (GLAM2). Además del descubrimiento de factores, MEME Suite proporciona herramientas para escanear secuencias de coincidencias con factores (FIMO, MAST y GLAM2Scan), escanear en busca de grupos de factores (MCAST), comparar factores con factores conocidos (Tomtom), encontrar espacios preferidos entre factores (SpaMo), prediciendo los roles biológicos de los factores (GOMo), midiendo el enriquecimiento posicional de secuencias para factores conocidos (CentriMo) y analizando ChIP-seq y otros conjuntos de datos grandes (MEME-ChIP). La figura 20 es un esquema de la interfaz descrita.

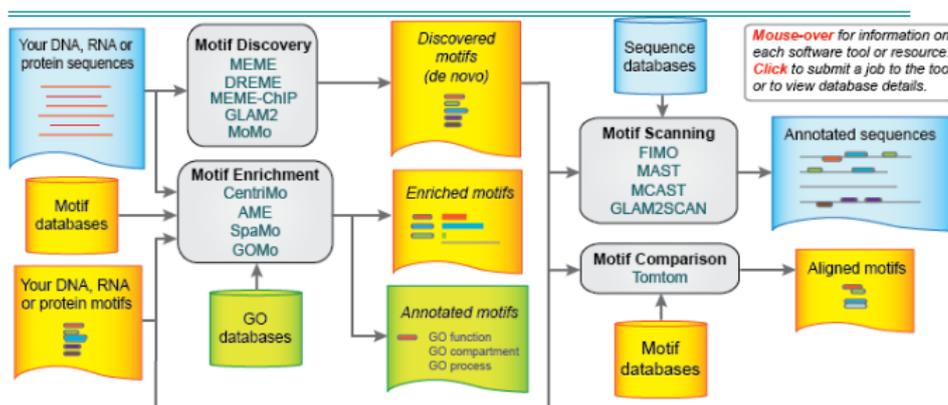


Figura 20: Esquema de la interfaz MEME Suite.

MEME Suite se compone de una colección de herramientas. La herramienta utilizada en este proyecto se encuentra en la sección de *Motif Scanning*, y se eligió FIMO (*Find Individual Motif Occurrences*) ya que es la que más se adecúa a lo que se busca. FIMO busca en una base de datos de secuencias dada las ocurrencias de factores conocidos. Este programa trata cada

⁷<http://meme-suite.org/index.html>

factor de forma independiente e informa todos los supuestos sucesos de factores por debajo de un umbral de valor-p especificado[94]. El resultado establece la cantidad de ocurrencias de factores encontrados y el valor-p de cada uno de ellos. Se menciona además que el valor-p de la aparición de un factor se define como la probabilidad de una secuencia aleatoria de la misma longitud que el factor que coincide con esa posición de la secuencia con una puntuación tan buena o mejor. La puntuación para la coincidencia de una posición en una secuencia con un factor se calcula sumando las entradas apropiadas de cada columna de la matriz de puntuación dependiente de la posición que representa el factor. El valor q de una ocurrencia de factor se define como la tasa de descubrimiento falso si la ocurrencia se acepta como significativa. Los resultados contienen una línea para cada coincidencia significativa con un factor. Las líneas se ordenan por orden de importancia estadística decreciente (valor-p creciente). La primera línea contiene los nombres (separados por tabulaciones) de los campos. Los nombres y significados de cada uno de los campos se describen en la tabla a continuación.

1. *Motif_{id}*: El nombre del factor, que es único en el archivo de base de datos de factores.
2. *Motif_{alt_id}*: Un nombre alternativo para el factor, que puede proporcionarse en el archivo de base de datos de factores.
3. *Sequence_{name}*: El identificador de la secuencia (de la línea de encabezado de secuencia FASTA) o el nombre de la secuencia extraída del identificador de secuencia.
4. Start: La posición de inicio de la aparición del factor; coordenadas de secuencia basadas en la secuencia ingresada.
5. Stop: La posición final de la aparición del factor; coordenadas de secuencia basadas en la secuencia ingresada.
6. Strand: Describe la hebra del ADN con la cual coincide el factor, '+' o '-'.
7. Puntaje: La puntuación para la aparición del factor. La puntuación para la coincidencia de una posición en una secuencia con un factor se calcula sumando la entrada apropiada de cada columna de la matriz de puntuación dependiente de la posición que representa el factor.
8. P-value: El valor-p de la aparición del factor. El valor-p de una coincidencia de factores es la probabilidad de una subsecuencia aleatoria única de la longitud de la puntuación de factores al menos tan bien como la coincidencia observada.
9. Q-value: El valor q de la aparición del factor. La tasa mínima de descubrimiento falso (FDR) requerida para considerar esta coincidencia significativa. FIMO estima los valores q de todos los valores p coincidentes utilizando el método propuesto por Benjamini y Hochberg[95].
10. Secuencia: La región de la secuencia coincide con el factor.

10.1.5. The Human Protein Atlas

The Human Protein Atlas (THPA) (<https://www.proteinatlas.org/>⁸) es un programa de origen sueco iniciado en 2003 con el fin de mapear todas las proteínas humanas en células, tejidos

⁸<https://www.proteinatlas.org/>

y órganos mediante la integración de diversas tecnologías. Es una plataforma de acceso abierto para permitir a los científicos, tanto académicos como industriales, acceder libremente a los datos para explorar el proteoma humano. El Atlas de proteínas humanas consta de seis partes separadas, cada una centrada en un aspecto particular del análisis de las proteínas humanas en todo el genoma; el Atlas de tejidos muestra la distribución de las proteínas en todos los tejidos y órganos principales del cuerpo humano, el Atlas de células muestra la localización subcelular de proteínas en células individuales, el Atlas de patología muestra el impacto de los niveles de proteínas para la supervivencia de los pacientes con cáncer; el Blood Atlas muestra si la proteína tiene una sobre-expresión o un rol determinado en las células sanguíneas; Brain Atlas explora la expresión de proteínas en el cerebro de los mamíferos mediante la integración de datos de tres especies de mamíferos (humanos, cerdos y ratones); y Metabolic Atlas permite la exploración de la función proteica y la expresión génica en el contexto de la red metabólica humana[96]. La figura 21 representa el crecimiento de la base de datos del atlas de proteínas, la cual se complementa con perfiles transcriptómicos de 27 tejidos normales, y un formato con seis sub-atlas.

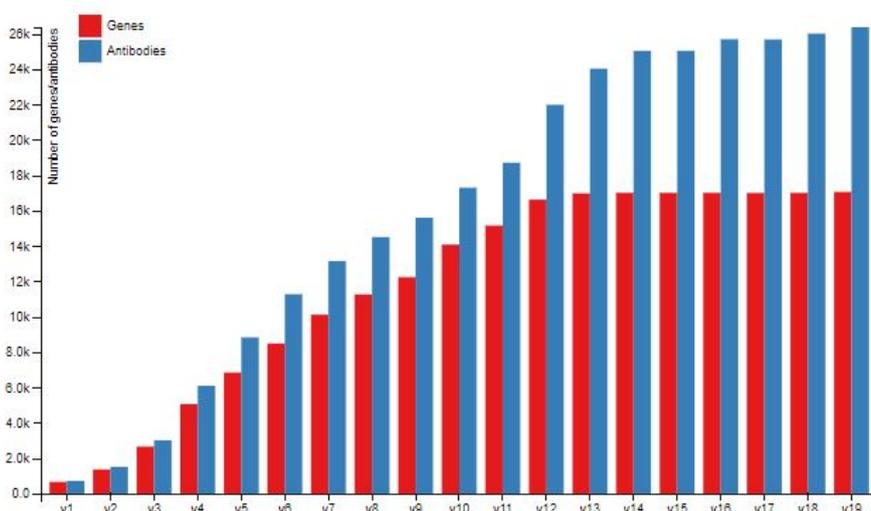


Figura 21: Figura representativa de la cantidad de genes y anticuerpos registrados en *The Human Protein Atlas* en el correr de los años.

El Atlas de proteínas humanas proporciona un recurso para la investigación biomédica basada en la patología, incluida la ciencia de las proteínas y el descubrimiento de biomarcadores. Además de contribuir a un mayor conocimiento básico sobre la enfermedad neoplásica, la identificación eventual de perfiles de proteínas específicas del tipo de cáncer tiene el potencial de mejorar la precisión del diagnóstico y formar una base para enfoques terapéuticos nuevos e individualizados[97].

10.1.6. Gepia

Gepia (Gene Expression Profiling Interactive Analysis) (<http://Gepia.cancer-pku.cn/>⁹) es una plataforma online que tiene funciones interactivas y personalizables que incluyen análisis

⁹<http://Gepia.cancer-pku.cn/>

de expresión diferencial, trazado de perfiles, análisis de correlación, análisis de supervivencia del paciente, detección de genes similares y análisis de reducción de dimensionalidad. Basada en datos TCGA y GTEx, Gepia permite el análisis de expresión completos basados en 9736 tumores y 8587 muestras normales de las bases de datos TCGA y GTEx. Los resultados del análisis abarcan 20 000 codificantes y 25 000 genes no codificantes, así como 14 000 pseudogenes y 400 segmentos de receptores de células T. Los parámetros de personalización flexibles de Gepia también permiten a los usuarios personalizar ampliamente la visualización, por ejemplo, cambiando los colores o modificando el ancho de los gráficos de expresión.

Gepia es una herramienta intuitiva y que ahorra tiempo para liberar el valor de los grandes datos genómicos en TCGA y GTEx. Gepia es una plataforma con el potencial de convertirse en una parte integral de los análisis de datos de rutina para los biólogos experimentales [98]. La figura 22 esquematiza los resultados de Gepia, y por lo tanto, toda la información aportada al usuario.

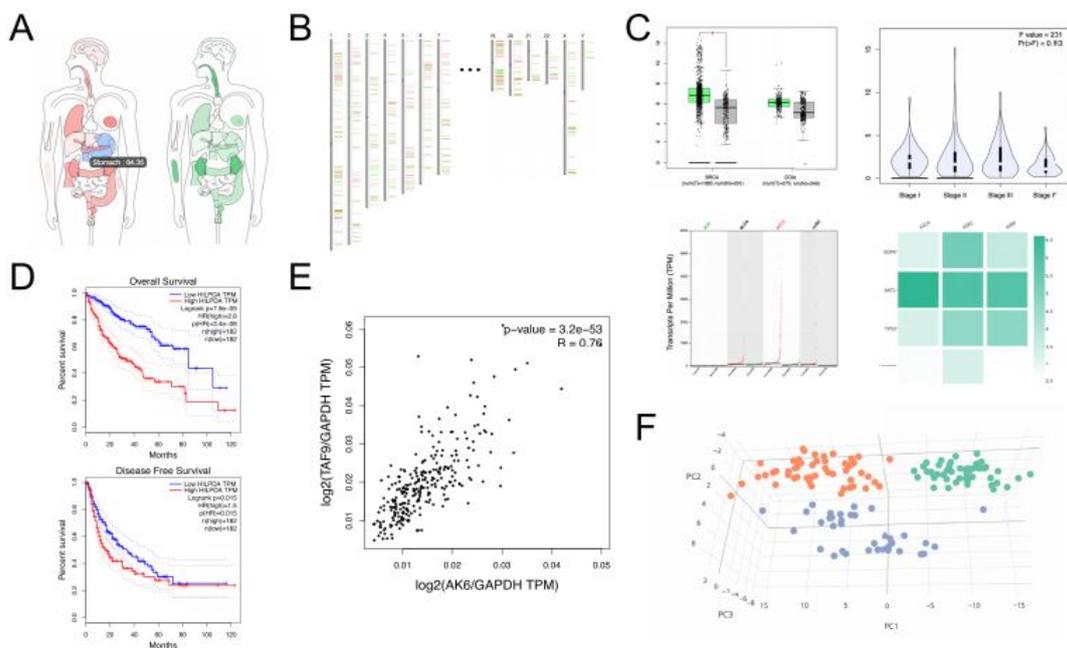


Figura 22: Resultados de Gepia. (A) Los usuarios pueden verificar la expresión génica por mapa corporal en la pestaña “General”. (B) Los genes diferenciales proporcionados por Gepia se pueden trazar como distribución cromosómica en la pestaña “Genes diferenciales”. (C) Gepia proporciona gráficos de caja, gráficos de violín basados en etapas patológicas, gráficos de puntos y gráficos de matriz. (D) El análisis de supervivencia general y supervivencia libre de enfermedad del gen de interés se puede presentar en la pestaña “Supervivencia”. Mientras tanto, se pueden identificar genes con la asociación más significativa con la supervivencia del paciente. (E) Gepia proporciona análisis de correlación de genes por pares para conjuntos de datos de expresión TCGA y/o GTEx. Los genes de interés pueden ser normalizados por otro gen. (F) Los resultados del análisis de componentes principales basados en un conjunto de genes de entrada se pueden presentar en la pestaña “PCA”.

10.1.7. SnapGene

SnapGene (<https://www.snapgene.com/>¹⁰) es una aplicación que permite planificar, visualizar y documentar los procedimientos de biología molecular. Es un software muy utilizado con el fin de construir modelos de genes y plásmidos en silico[99][100].

11. MÉTODOS

11.1. Búsqueda de los factores de transcripción

Primero, se trabajó con las plataformas online *MATCH*, *PROMO* y *MEME SUITE - FIMO* para obtener los TFs de los ocho promotores seleccionados. Un factor importante a tener en cuenta es que la estrategia de filtrado elegida determina el criterio con el que se deben analizar los resultados. Es decir, en caso de elegir minimizar los falsos positivos, se busca un tipo de prueba altamente específica, los resultados serán de unos pocos factores comparando con minimizar falsos negativos. Como bien se mencionó antes, puede ocurrir que si bien los valores de probabilidad son altos, los mismos sean falsos positivos, y es por eso que se debe analizar los resultados obtenidos en función de la estrategia elegida. A su vez, los valores de probabilidad sirve para filtrar de entre los resultados los que más se asemejan a la secuencia brindada por el usuario. La metodología utilizada para obtener los resultados para los ocho promotores se explica a continuación.

1. *MATCH*: Se obtuvieron los resultados con el filtro para minimizar la suma de ambas tasas de error, tanto los falsos positivos como los falsos negativos. Se eligió con el fin de tener un equilibrio entre la especificidad y sensibilidad de la prueba. Luego se obtuvieron los resultados con el filtro para minimizar los falsos positivos, con el fin de tener la máxima sensibilidad posible. Los TFs que se obtuvieron se incluyeron en los factores obtenidos en la tabla inicial, a pesar de que los valores de probabilidad no necesariamente fuesen buenos. De esta manera, se busca incluir todos los verdaderos positivos.
2. *PROMO*: Se obtuvieron los resultados con el valor-predeterminado de 15 % (85 % de similitud)
3. *FIMO*: Una vez realizados los pasos explicados en la sección *Materiales*, se obtuvieron los resultados.

Una vez obtenidas las tablas, para las tres plataformas online se corrigió la posición de los TFs encontrados en función del inicio de transcripción del promotor. Para realizar esto se utilizó el programa estadístico RStudio (versión 3.5.7), donde se importaron las tablas y se formatearon. Luego, se realizó un primer filtro eliminando los TFs encontrados *río abajo* del inicio de transcripción.

¹⁰<https://www.snapgene.com/>

11.2. Terminología

Antes de realizar el análisis con todos los factores de transcripción encontrados, se unificó la nomenclatura de los TFs de las distintas plataformas online. Esto era necesario ya que cada plataforma tiene su propia terminología y en muchos casos, se encontraba el mismo TF en la misma posición en dos bases de datos distintas pero con distintos nombres. Por lo tanto, todos los TFs se los llamó con el símbolo del Comité de Nomenclatura Genética de HUGO (HGNC) correspondiente para poder así comparar los resultados obtenidos. HGNC es un comité de la Organización del Genoma Humano (HUGO) que establece los estándares para la nomenclatura de genes humanos. HGNC aprueba un nombre único y significativo para cada gen humano conocido, basado en una consulta de expertos [101]. Además, esto permite que los TFs sean considerados como válidos para *ToppGene* que se utiliza en la priorización génica.

11.3. Priorización génica

Una vez obtenidas las tablas de las plataformas online y modificadas según el proceso descrito en *Búsqueda de los factores de transcripción*, se utilizó la herramienta *ToppGene* para el análisis de la priorización génica. Por lo tanto, como lista de entrenamiento se utilizaron los promotores seleccionados, y como lista de prueba se utilizaron todos los TFs obtenidos en las diferentes plataformas.

11.4. Análisis de datos

11.4.1. Factores de transcripción

Antes de continuar explicando el proceso de análisis, se debe mencionar que en este trabajo se decidió trabajar con las tres plataformas online presentadas en la sección de *Materiales*, y se utilizó la herramienta de *ToppGene* para el análisis de la priorización génica, tal como se describió previamente. En lugar de optar por un análisis con los valores estadísticos ya brindados por cada una de las plataformas online, se utilizó otra herramienta ajena a estos softwares ya que *ToppGene* no solo es una herramienta más específica en cuanto a la información que devuelve, con un valor-p ajustado para los datos en función de las catorce medidas de similitud (como fue mencionado en la sección de *Materiales*), sino que al utilizar tres plataformas online que implementan distintas bases de datos con algoritmos propios e independientes del resto, determinar un único punto de corte para tomar los datos como válidos no era viable para las tres plataformas. De esta manera, utilizando *ToppGene*, se puede tomar como válido un único punto de corte de valor-p igual a 0.05. Los datos que tienen un valor-p menor al punto de corte se consideran estadísticamente significativos. Para el caso particular de VWA2, la plataforma *ToppGene* no tiene resultados estadísticamente significativos para dicho gen. Como forma parte de los objetivos analizar dicho promotor ya que es de interés para el Instituto Leloir, se realizó un método particular. Para el mismo, se trabajó directamente con las plataformas online mencionadas en *Materiales*, de la siguiente manera:

- Primero se estudió con RStudio cuales eran los TFs que se encontraban en común en dos o tres bases de datos.

- Luego se incluyeron los TFs que eran considerados estadísticamente significativos para cada base de datos. En particular, para *MATCH* se hizo un análisis y se incluyó los TFs resultantes de aplicar el filtro de “minimizar falsos positivos”. Por otro lado, para *PROMO*, se verificó aquellos factores que tenían un valor de disimilitud bajo (cercano a 0). Finalmente, para *FIMO* se incluyeron los factores que tenían valor-p menor a la media de los valores p de los resultados.

Como resultado, ya sea para los promotores A33, AQP5, CDC25A, CDC25B, GLI1, FSCN1, NQO1, cuyo análisis incluyó la herramienta *ToppGene*, como para VWA2, cuyo proceso de análisis se basó únicamente en las plataformas, las tablas finales incluyen los factores de transcripción más estadísticamente significativos.

11.4.2. Posiciones

Una vez obtenidos los TFs para los ocho promotores, el análisis de las posiciones para cada uno de ellos se basó en el estudio de las posiciones para cada plataforma independientemente. Por lo tanto, los resultados finales de las posiciones son aquellas que tienen valor de probabilidad más estadísticamente significativa para cada plataforma. Es decir, para *MATCH* se hizo un análisis y se incluyó las posiciones con *core match* y *query match* cercanos a 1. Por otro lado, para *PROMO*, se verificó aquellas posiciones que tenían un valor de *dissimilitud* bajo (cercano a 0). Finalmente, para *FIMO* se incluyeron las posiciones que tenían *valor-p* menor a la media de los valores p de los resultados.

11.4.3. Definición de variables para promotores híbridos

En este proyecto, las variables de estudio fueron aquellas que otorgan especial interés a los factores de transcripción que conforman al promotor como específico de tumor. Las variables se describen en la tabla.

Definición de variables		
Variable	Tipo	Valores
Especificidad de cáncer	Categoría dicotómica	!X: sin rol determinado en la enfermedad de cáncer X: con rol determinado en la enfermedad de cáncer
Especificidad de tejido	Categoría dicotómica	!X: sin sobre-expresión en el tejido intestinal X: con sobre-expresión en el tejido intestinal

Cuadro 2: Definición de variables para promotores híbridos.

- **Especificidad de cáncer:** Se busca identificar si el TF tiene un rol determinado en la enfermedad. **!X** significa que en la actualidad no hay un rol definido para el mismo, pero puede ocurrir que diversos papers busquen y comprueben roles específicos en un tipo de cáncer que aún faltan más estudios para corroborarlo. **X** significa que actualmente está comprobado que el TF tiene un rol determinado en la enfermedad. Teniendo en cuenta

que en el cáncer se activan redes reguladoras de la transcripción que no se encuentran en las células normales, detectar los TFs involucrados en estas redes y la función que cumplen nos permite diseñar promotores tanto híbridos como sintéticos que solo deberían estar activos en las células cancerosas. **!X** significa que actualmente no está comprobado que el TF tiene un rol determinado en la enfermedad.

- **Especificidad de tejido:** Se busca identificar, como bien dice la tabla, si hay sobre-expresión en el tejido intestinal. Sin embargo, **!X** no quiere decir que no haya expresión en dicho tejido, quiere decir que no hay una sobre-expresión por sobre el resto de los tejidos. **X** quiere decir que efectivamente el factor se encuentra sobre-expresado en dicho tejido.

11.4.4. Definición de variables para promotores sintéticos

En este proyecto, las variables de estudio fueron aquellas que otorgan especial interés a los TFs que conforman al promotor sintético como específico de tumor. Las variables se describen en la tabla 3.

Definición de variables		
Variable	Tipo	Valores
Especificidad de cáncer	Categórica dicotómica	!X: sin rol determinado en la enfermedad de cáncer X: con rol determinado en la enfermedad de cáncer
Especificidad de tejido	Categórica dicotómica	!X: sin sobre-expresión en el tejido intestinal X: con sobre-expresión en el tejido intestinal
Actividad viral	Categórica dicotómica	!X: sin rol determinado en la actividad viral X: con rol determinado en la actividad viral
Actividad en respuesta inmunológica	Categórica dicotómica	!X: sin rol determinado en la actividad de rta. inmunológica del organismo X: con rol determinado en la actividad de rta. inmunológica del organismo

Cuadro 3: Definición de variables para promotores sintéticos.

Teniendo en cuenta que las dos primeras fueron explicadas anteriormente, se describen únicamente las dos últimas.

- **Actividad viral:** Se busca una relación entre el TF y la actividad viral, teniendo en cuenta que todo el análisis de promotores es para modificar genéticamente un adenovirus y optimizar su performance en las células necróticas de tejido colorrectal.
- **Actividad en respuesta inmunológica:** Se busca identificar si existe una relación entre el TF y la respuesta inmunológica al organismo ya que es una característica elemental que aporta información a los investigadores del *Instituto Leloir*, tratándose este proyecto de una inmunoviroterapia.

Es importante aclarar que las variables explicadas recientemente se incorporan al análisis para el caso de promotores sintéticos ya que sin ellas, muchos TFs relevantes en el proceso de la enfermedad serían dejados fuera del estudio. Teniendo en cuenta que el armado de promotores sintéticos se compone de componentes modulares correspondientes a los factores, se debe tener en cuenta a aquellos que tienen actividad viral ya que interactúan con los genes del virus en sí, o tienen un rol en la respuesta inflamatoria del organismo al cual se aplica la terapia. Por lo tanto, dado que se tienen que tener en cuenta más características para el armado, se incorporan las variables definidas respectivamente. Con respecto a los **promotores híbridos**, se analizan los TFs según las variables **especificidad de cáncer** y **especificidad de tejido** ya que el interés de estudio es entender cuál de los presentados por el laboratorio tiene TFs que expliquen la expresión diferencial de la proteína.

11.4.5. *Data Mining* para los factores de transcripción resultantes

En este proyecto, se consideraron válidas las bases de datos para el análisis pertinente de los TFs de *THPA*, *Gepia*, *ToppFun*, *Gene Ontology*, *KEGG/WikiPathways* y el repositorio público de papers publicados en *Pubmed*, remitidos por la comunidad científica. Principalmente, se utilizaron *THPA* y *Gepia* para analizar la sobre-expresión de las proteínas en tejidos y si el factor analizado tiene un rol determinado en la enfermedad de *cáncer* y por lo tanto se lo conoce como un *cancer-related gene* o *proto-oncogene*. Se buscó específicamente para cada TF, si tenían alta afinidad con un tejido, cumplían un rol fundamental con el cáncer, si tenían un rol determinado en la actividad viral y si tenían un rol determinado en la actividad inmunológica. A su vez, para cada factor estadísticamente significativo se buscó el rol específico y posible interacción con otros factores presentes en los resultados.

11.4.6. Armado de los promotores híbridos con los respectivos factores de transcripción

Una vez obtenidos los TFs con las posiciones correspondientes, se diseñó la secuencia del promotor *in silico* utilizando el software SnapGene descrito en la sección *Materiales*. Por lo tanto, para realizar los modelos finales de promotores con los factores se implementó el programa SnapGene ya que permite una clara visualización de los diseños finales. La figura 23 esquematiza la metodología descrita hasta el momento para los promotores seleccionados. Este análisis es para los promotores híbridos.

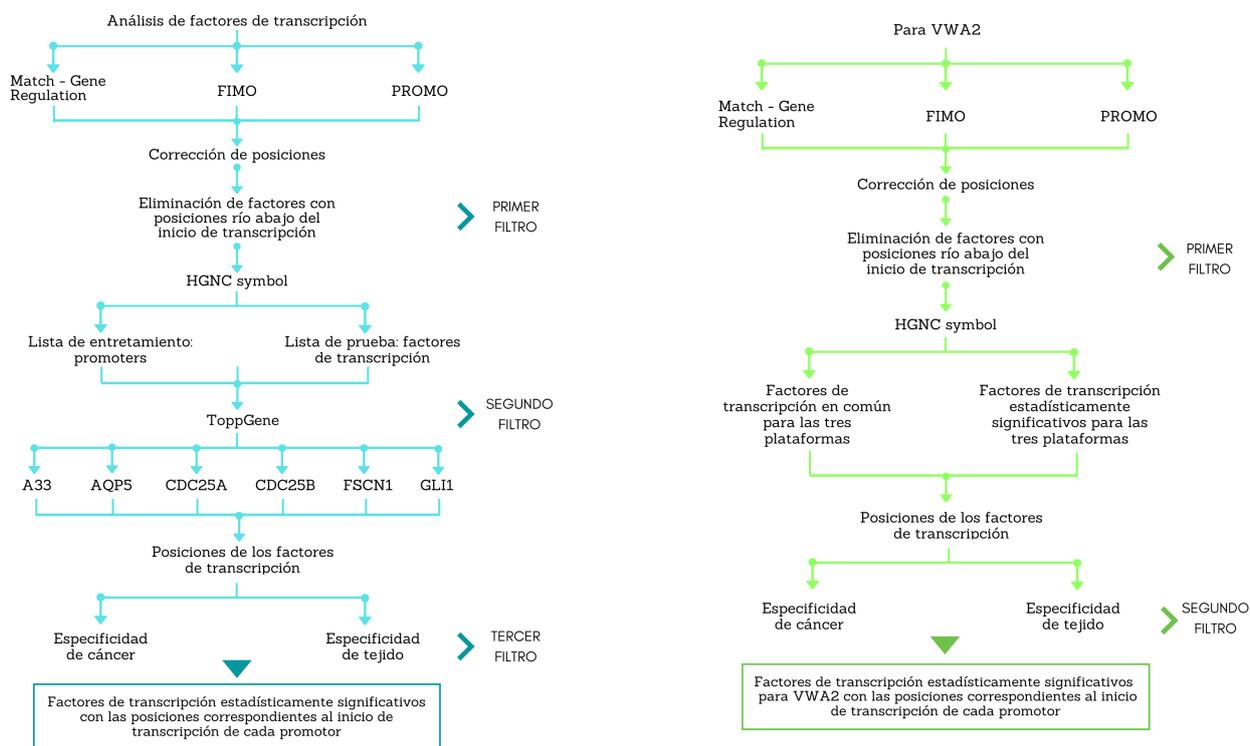


Figura 23: Diagrama de flujo de la metodología empleada para los promotores híbridos. A la izquierda es el proceso para A33, AQP5, CDC25A, CDC25B, FSCN1, GLI1 y NQO1 y a la derecha para VWA2.

11.4.7. Selección de factores de transcripción para el armado de promotores sintéticos

Para el caso de los promotores sintéticos, la metodología empleada es muy similar. La única diferencia es que se analizan los factores finales en función de las variables presentadas en la tabla 3. La figura 24 esquematiza la metodología descrita hasta el momento para la selección de factores en el armado de promotores sintéticos.

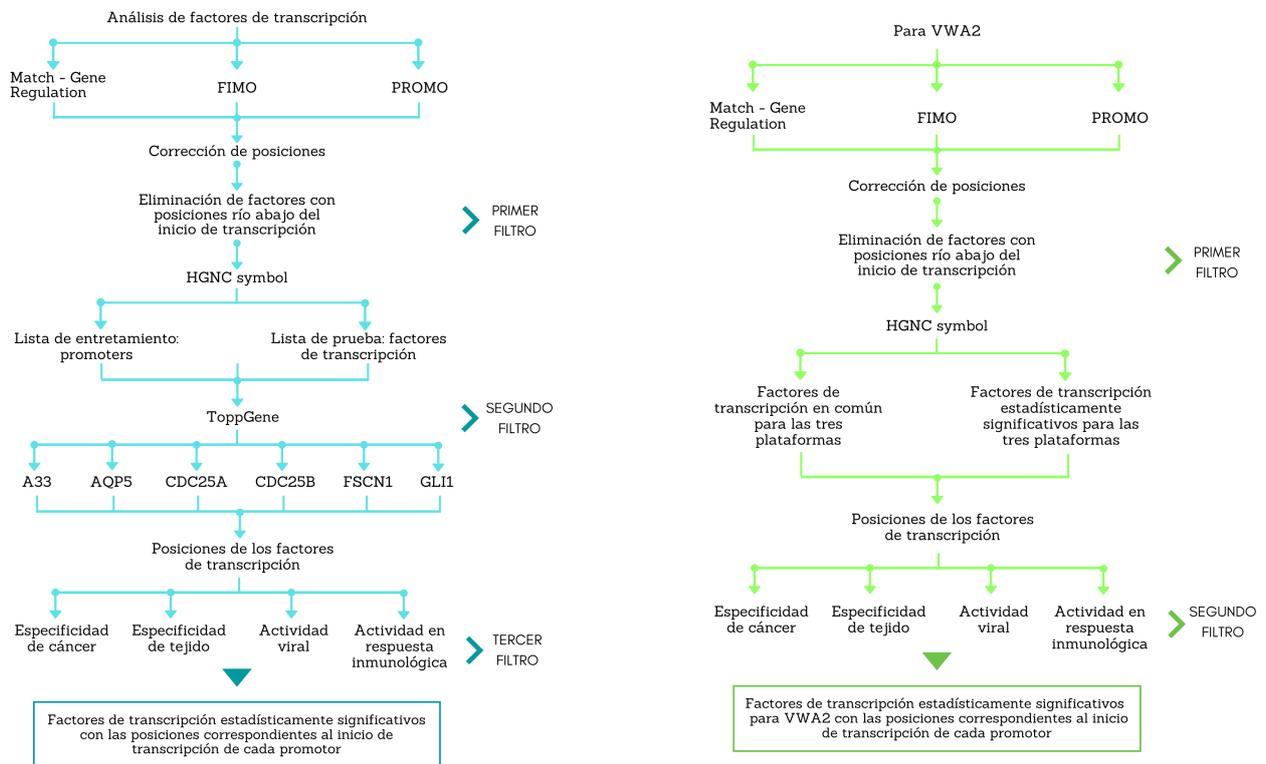


Figura 24: Diagrama de flujo de la metodología empleada para los promotores sintéticos. A la izquierda es el proceso para A33, AQP5, CDC25A, CDC25B, FSCN1, GLI1 y NQO1 y a la derecha para VWA2.

Resultados

12. Análisis de los factores de transcripción de los 8 promotores y las distancias al inicio de transcripción

12.1. Primer análisis de los promotores

Un primer análisis se basó en el estudio de los ocho promotores. El mismo consistió en un proceso de *Minería de texto* con la información actual relacionada a los TFs de cada promotor conocidos hasta el momento. La tabla 4 muestra a modo resumen los resultados obtenidos.

Promotores	Factores de transcripción
A33	CDX1, CDX2, KLF4, KLF5 [31][30]
AQP5	NF-KB[102]
CDC25A	c-myc, E2F, TGFB1 [106][115], SP1, AP-2, factores CCAAT(CEBPA, CEBPB, CEBPD), b-myb [113], STAT3[45]
CDC25B	TP53, SP1, NFYB [116]
FSCN1	CREB, AhR [51]
GLI1	CCND2-Cyclin D2, FOXM1, CD133, SOX9, NFkB [55][117][58]
NQO1	AP-1, AP-2, XREs, AREs (NFE2L2, NFE2L3), c-Jun[118][60], tp53, c-myc [59]
VWA2	c-myc[63]

Cuadro 4: Resultados de estudios previos que asocian factores de transcripción a los promotores estudiados.

A33: El grupo de investigación de (Cameron N. Johnstone et al; 2002) sostienen que la actividad de CDX1 y CDX2 es fundamental para mediar el patrón de expresión específico del intestino del gen del antígeno A33[31]. El grupo sostiene que A33 es uno de los pocos que presenta el factor homeobox específico del intestino, CDX1 y CDX2. Esto explicaría la alta especificidad y potencia del promotor. A su vez, los investigadores mantienen que la alineación de la región promotora reveló varios sitios de unión conservados para TFs previamente implicados en la regulación de la expresión génica en células epiteliales intestinales. Entre ellos se encuentran los TFs de la familia Kruppel (GKLF / KLF4 e IKLF / KLF5) y dos supuestos sitios de unión a CDX. A su vez, el grupo de (Zebin Mao et al; 2003) sostiene que GKLF juega un papel crítico

en la decisión entre la proliferación y la detención / diferenciación del ciclo celular y sugieren que GKLf es un TF pleiotrópico que regula la expresión del antígeno A33, funcionando como activador en las células de carcinoma de colon [30].

AQP5: Se sabe que el proceso EMT está mediado por una serie de vías de señalización, incluidas las vías de señalización de factor nuclear kappa B (NF- κ B), factor de crecimiento transformante b (TGF- β), Wnt y Notch. A su vez, se debe tener en cuenta que el silenciamiento de AQP5 inhibe la migración y la invasión de las células de CCR, lo que se acompaña de una expresión alterada de las moléculas relacionadas con EMT y una transducción de señalización debilitada de Wnt / b-catenina. Los cambios fenotípicos celulares inducidos por el silenciamiento de AQP5 se restauran parcialmente cuando se reactiva la vía Wnt / b-catenina [34][102][103]. Esto es importante ya que puede explicar varios de los TFs del promotor, entre ellos NF- κ B1, AP-1, PPAR, Cyclin-D1, MYC, entre otros.

CDC25A: Teniendo en cuenta los resultados del grupo de investigación (Xundi Xu et al; 2003), si bien el mecanismo para la sobreexpresión de CDC25A en cáncer es desconocido, algunas moléculas como c-myc, E2F y TGFB1 podrían considerarse como candidatos clave para afectar la síntesis de CDC25A[106]. A su vez, el grupo de (Luca Busino et al; 2004) coincide en la presencia de la activación transcripcional mediada por E2F-1- y c-Myc en dicho promotor [115]. Por otro lado, el trabajo de (Antonio Iavarone et al; 1999) muestra que el elemento de respuesta inhibitoria de TGF- β en el promotor CDC25A se asigna a un sitio E2F en los nucleótidos -62 a -55 desde el sitio de inicio de la transcripción. A su vez, sostiene que los TFs AP-2, SP1, factores CCAAT, y b-myb también se encuentran en la secuencia del promotor[113]. Por último, STAT3 actúa como un enhancer en el promotor de CDC25A, y esto se corrobora en el estudio de Yanchao Ma et al., 2020, donde se inhibió la fosforilación de STAT3 en las células HCT116 (línea celular de cáncer de colon humano utilizada en investigación terapéutica y exámenes de drogas) y RKO (es una línea celular de carcinoma de colon) y como consecuencia, disminuyó significativamente la expresión de CDC25A [45].

CDC25B: Según el grupo de investigación de (M Dalvai et al; 2003), los mismos sostienen que en presencia de TP53 se activa el promotor CDC25B, evitando la unión de Sp1 y aumentando la unión de NFYB en el promotor CDC25B. A su vez, el grupo sostiene que la región -122 a -51 del promotor contiene dos factores ricos en GC para unir Sp1 / Sp3, mostrados como importantes para la actividad transcripcional del promotor CDC25B [116].

FSCN1: Para el grupo de investigación de (Yosuke Hashimoto et al; 2009), los hallazgos demuestran roles previamente no identificados de los TFs AMPc (CREB) y el receptor de hidrocarburos de arilo (AhR) en la regulación de la transcripción de FSCN1 en células de carcinoma humano. Estos datos identifican los factores candidatos AhR y CREB como determinantes principales de la actividad transcripcional de la región FSCN1 -219/+7. Estos resultados demuestran una correlación entre la asociación de CREB y AhR con la región -219/+7 del promotor FSCN1 y la expresión de la proteína fascina en células de carcinoma de colon humano[51].

GLI1: Se debe tener en cuenta que los resultados del grupo de investigación de (Yin-lu Ding et al; 2012) sostienen no solo que las evidencias acumuladas han demostrado que GLI1 está involucrado en la progresión tumoral, sino además que la ciclina D2 y FOXM1 se han identificado como los genes diana aguas arriba de GLI1 en los carcinomas de células basales[55]. Años más tarde, el grupo de estudio de (Chuan Zhang et al; 2016) reafirman que FOXM1 desempeña un papel importante en la aparición y el desarrollo de tumores malignos y regula la progresión

tumoral como un TF aguas arriba de Gli1 en cánceres como los carcinomas de células basales y los carcinomas de pulmón el nivel. Por lo tanto, concluyen que FOXM1 tiene una correlación positiva con GLI1, y a su vez, que GLI1 promueve la metástasis de las células de CCR de una manera dependiente de FOXM1 activando la señalización EMT y PI3K-AKT[117]. A su vez, el grupo de (Zhaoting Yang et al; 2018) mostraron que la expresión de GLI1 se correlacionó significativamente con la expresión de genes relacionados con la madurez como CD133 y SOX9, pero no con otros marcadores de células madres cancerosas (CSC) como LSD1, SOX2, CD44 y LGR5. A su vez, el mismo grupo sostiene la población de células positivas para GLI1 tiene una gran superposición con la población de células positivas para NFkB p65[58].

NQO1: La investigadora Ani K. Jaiswal en 1991 identificó la secuencia de nucleótidos de un elemento de respuesta antioxidante ARE caracterizada en el gen NQO1 humano. Curiosamente, la secuencia de consenso para la unión a la proteína AP1 (TGACTCA) está contenida dentro de la secuencia ARE (TCACAGTGACTCAGCAGAATC) del gen NQO1 humano. A su vez, identifican secuencias asociadas a elementos de respuesta xenobiótica (XREs), entre los cuales se asocian los TFs NR1I2, RORA, NR1F1, ARNT, entre otros. Asimismo, identifican secuencias asociadas a los factores de AP-2, secuencias de cajas TATA y CCAAT, y elementos de respuesta antioxidante (AREs) tales como NFE2L2 (también conocido como Nrf2) y NFE2L3 (también conocido como Nrf3) y AR (Androgen receptor)[60][118]. El complejo de TF ARE es necesario tanto para la inducción como para la expresión de NQO1 y otros genes de enzimas desintoxicantes. En este proceso, Nrf2 con c-Jun se une al ARE y activa la expresión génica [60]. A su vez, el grupo de (Lili Ji et al; 2014) sostiene que Nrf2, NQO1, MRP1, c-myc y TP53 se expresaron altamente en tejidos de CCR en comparación con tejidos adyacentes no tumorales, y hay correlaciones positivas significativas entre la expresión de Nrf2 y la de NQO1, MRP1, c-myc y TP53[59].

VWA2: El grupo de investigación de Beatriz González et al., descubrió en 2018 que la expresión de VWA2 en CCR primarios se correlaciona con la de los genes Myc y Myc-target[63]. Actualmente, no se encuentran mayores resultados que asocian TFs al promotor.

Este proceso se realizó para luego poder comparar y validar los resultados de los TFs obtenidos *in silico*. En el mejor de los casos, los factores estudiados hasta el momento deberían aparecer en la lista de TFs finales presentados.

12.2. Búsqueda de los factores de transcripción

Con el objetivo de identificar los TFs de los respectivos promotores, se realizaron los pasos descritos en la sección de *Métodos*. La figura 23 y 24 esquematiza la metodología empleada para los promotores híbridos y sintéticos con A33, AQP5, CDC25A, CDC25B, FSCN1, GLI1 y NQO1 y VWA2, es decir, aquellos que obtuvieron resultados estadísticamente significativos en la priorización génica de ToppGene y VWA2 que no lo tuvo.

12.3. Terminología

Una vez obtenidas las tablas, y habiendo realizado el primer filtro, se agregó una nueva columna con a terminología de cada TF correspondiente al símbolo de HGNC, unificando de esta manera la nomenclatura de los mismos.

12.4. Posiciones

Para definir las posiciones de los factores encontrados, se estructuró a partir de las plataformas online de manera independiente. Es decir, se buscó las posiciones de las tres plataformas de manera individual, y una vez obtenidas, se comparó las posiciones encontradas para un mismo factor. En el caso en el que ya existiese una posición estudiada para ese factor para ese promotor en particular, se priorizaba dicha posición. En la gran mayoría de los casos, los factores ya estudiados se encontraban en dos o tres plataformas con posiciones coincidentes. Esto demuestra la robustez de ciertos resultados, ya que no solo eran encontrados por tres plataformas que tienen algoritmos de búsqueda distintos, sino que eran coincidentes en las posiciones. En el caso en el cual no se conocía la posición, y el factor era encontrado por una única base de datos, se buscaban las posiciones con mejor resultado en cuanto al valor de probabilidad. Este depende de cada plataforma, como se vio anteriormente, ya que en el caso de PROMO, los mismos se evalúan en función de **Dissimilarity**, **RE equally** y **RE query**. Por otro lado, para FIMO, los valores se informan con el **valor-p** y el **valor q** de los resultados. Por último, para MATCH se analizan los valores relacionados a **core match** y **matrix match**. Los mismos se encuentran explicados en la sección de *Materiales*. En el caso en el cual dos o tres plataformas presentaban posiciones para el mismo factor, se priorizaba las posiciones de FIMO y MATCH frente a las de PROMO. Esto es así ya que PROMO presentaba muchas posiciones con valores de probabilidad muy buenos ya que la longitud del factor encontrado era de pocas pares de base (2 o 3) frente al mismo factor encontrado en las otras dos plataformas con longitudes más largas. Un claro ejemplo de esta degeneración de posibles resultados se presenta en la tabla 10 de factores de GLI1, con el factor CEBPA. El mismo se encuentra en las tres plataformas, FIMO y MATCH presentan una única posición del mismo (que además coincide para ambas plataformas) con longitudes entre 11 a 13 pares de bases, mientras que en PROMO se encuentran 6 posiciones posibles con longitudes de 5 pares de bases.

12.5. Proceso de *Minería de texto* con los factores de transcripción determinados

A través de los resultados obtenidos por el análisis de factores y la priorización génica con respecto a cada uno de las proteínas de interés, se realiza un proceso de minería de texto a través de las herramientas indicadas en la sección de **Materiales**. Por lo tanto, considerando todos los factores de todos los promotores identificados como estadísticamente significativos luego del ranking genético, se categorizan las variables en función de las variables identificadas en la tabla 3. Por lo tanto, para la sección de promotores híbridos se analizan solo los resultados de dos primeras columnas, mientras que para promotores sintéticos todas las columnas de la tabla son de interés. La tabla 13 contiene los resultados correspondientes. A su vez, la tabla 14, presenta una descripción más detallada de la función de los TFs. Esto se realizó para una mayor comprensión de los resultados, ya que muchos TFs no tienen ninguna de las características determinadas por las variables definidas, y sin embargo al leer la descripción de los procesos en los cuales se encuentra regulando, se entiende el por qué de su aparición, no solo como subunidad necesaria para activar a otros factores, sino la intervención del mismo en procesos que aún se encuentran en fase de investigación. Es por esto que, si bien las bases de datos de **The Human Protein Atlas** y **Gepia** se consideran más relevantes, se tiene en cuenta las

investigaciones actuales acerca de las vías de regularización de dichos factores que se encuentran en pleno desarrollo. También hay casos de estudios donde se dio a conocer un rol de un factor en un caos clínico en particular. Se priorizó la búsqueda de información en CCR. Por lo tanto, dado que la información de cada TF es extensa, se hizo un resumen de lo que se conoce en la enfermedad seleccionada y si su actividad influye en algún factor determinado. Esto es lo que se presenta en la tabla 14. A su vez, se incluyó la información relacionada a la sobre-expresión de tejidos distintos al intestinal ya que se considera relevante a la hora de hacer el análisis de cada uno.

12.6. Resultado de los factores de transcripción de cada uno de los promotores y las distancias al inicio

A continuación, las tablas presentan los resultados de los 8 promotores con los factores correspondientes. Para A33, AQP5, CDC25A, CDC25B, FSCN1, GLI1 y NQO1 se presenta el valor-p ajustado de *ToppGene*, junto con las posiciones finales para cada plataforma. Para VWA2 no se presenta el valor-p ajustado ya que la metodología empleada para obtener los resultados fue distinta.

Factor de transcripción	ValorP ajustado	PROMO	FIMO	MATCH
KLF4	0,0014766027		"-91, -81"	
CDX2	0,0016444091		"-40, -30"	
TP53	0,0022258365	"-95, -89"		
FLI1	0,0065867341		"-30, -21"	
MYB	0,0094754420	"-23, -16"		"-22, -12"
FOXP3	0,0424842733	"-94, -89; -19, -14"		

Cuadro 5: Factores de transcripción para el promotor A33 con las posiciones correspondientes al inicio de transcripción y el valor-p ajustado de *ToppGene*.

Factor de transcripción	ValorP ajustado	PROMO	FIMO	MATCH
STAT3	0,0001036860		"-221, -211"	
NFKB1	0,0001548420	"-340, -329; -339, -330"	"-340, -328"	
NR3C1	0,0002066212	"-341, -331"		
TP53	0,0002255620	"-246, -240; -201, -195; -131, -125; -119, -113"		
STAT1	0,0002344131	"-223, -214; -14, -5"		

Continúa en la siguiente página

Cuadro 6 (AQP5) – continuación de la página anterior

Factor de transcripción	ValorP ajustado	PROMO	FIMO	MATCH
SOX2	0,0003328267	“-346, -340”		
FOXA1	0,0004963064	“-225, -220”		
E2F1	0,0009749675	“-114, -108; -75, -69”		
RXRA	0,0010794233	“-395, -387; -184, -174; -136, -126; -57, -47”		
NF1	0,0013372724	“-327, -322; -239, -234; -212, -207”		“-250, -221”
WT1	0,0014793766	“-118, -109; -117, -107”		
JUN	0,0014854998	“-35, -30”		
ESR1	0,0015448447	“-36, -29”	“-247, -231; -35, -19”	“-36, -17”
NKX2-1	0,0018226019	“-314, -308”		
RELA	0,0019776562		“-338, -329”	
PAX6	0,0024287679	“-358, -354; -353, -349; -136, -132; -72, -68; -33, -29; -19, -15”		
HMGB1	0,0025989315	“-224, -218; -10, -4”		
NFE2L2	0,0027793528	“-32, -26”		
THRB	0,0027905824	“-37, -29”		
RARG	0,0028690298	“-92, -85”		
SMAD1	0,0029034345	“-87, -76”		
FOS	0,0029794869	“-37, -30; -36, -30”		
CEBPA	0,0032172875	“-302, -297; -300, -295; -261, -256; -259, -254”		
EGR1	0,0032405995		“-118, -105”	
HES1	0,0037037287	“-235, -229; -162, -156”		
VDR	0,0042136892	“-335, -332; -133, -130; -54, -51; -44, -41; -35, -32”		
MSX1	0,0046866096	“-235, -233; -231, -229; -226, -224; -162, -160; -100, -98; -91, -89; -61, -59; -42, -40”		
MYOG	0,0061170989	“-279, -273; -243, -237; -51, -45; -1, 5”		“-250, -221”
STAT5A	0,0062616003	“-338, -335; -221, -218; -204, -201; -10, -7”		
ZEB1	0,0062948192	“-290, -283; -81, -74”		
CEBPB	0,0068803555	“-300, -295; -259, -254”		
ESR2	0,0070537073	“-38, -30”	“-246, -232; -34, -20”	

Continúa en la siguiente página

Cuadro 6 (AQP5) – continuación de la página anterior

Factor de transcripción	ValorP ajustado	PROMO	FIMO	MATCH
RARB	0,0075201018	"-37, -30"		
YY1	0,0084048255	"-350, -342"		
NR1I2	0,0098401809	"-295, -287"		
THRA	0,0101050857	"-38, -30"		
SP1	0,0113621828		"-120,-110 ; -6, 4"	
PAX9	0,0121113688	"-235, -229; -104, -98; -46, -40"		
ASCL1	0,0128350579		"-3, 9"	
TFAP2A	0,0139887117	"-329, -324; -315, -310; -244, -239; -179, -174; -41, -36"	"-315, -304; -110, -100"	
CUX1	0,0143358388	"-260, -255"		
NR2F2	0,0178091284	"-38, -30"		
ETS1	0,0190273297	"-316, -311; -40, -35"		"-315, -305; -31, -21"
SP3	0,0200898197		"-120, -110; -45, -35"	
MTUS1	0,0202482801	" -287, -278"		
PPARA	0,0220294498	"-36, -30"		
EP300	0,0243917228	"-294, -290; -262, -258; -60, -56"		
SF1	0,0244997812	"-343, -335; -201, -193"		
IKZF1	0,0253620109	329, -321; -221, -213"		224, -211"
PAX2	0,0255070280	" -236, -230"		
MYOD1	0,0263899203	" -51, -44; -2, 5"		
JUND	0,0265300311	"-36, -30"		
E2F4	0,0278522839		"-116, -106"	
JUNB	0,0322024369	"-35, -30"		
ETV4	0,0332076108	"-29, -22; -13, -6"		
IRF1	0,0341927463	"-222, -214; -14, -6"		
NKX2-5	0,0344415118	"-226, -218"		
STAT6	0,0393830985	"-221, -213"		
ITGAM	0,0420777960	" -12, -6"		
HNF1B	0,0478325231	"-340, -331"		
ATF3	0,0492286966	"-37, -30"		

Cuadro 6: Factores de transcripción para el promotor AQP5 con las posiciones correspondientes al inicio de transcripción y el valor-p ajustado de ToppGene.

Factor de transcripción	ValorP ajustado	PROMO	FIMO	MATCH
TP53	9,2432950182797E-11	"-385, -379; -301, -295; -40, -34; -26, -20"		
E2F1	3,30088308908572E-08	"-158, -152; -61, -53"	"-61, -50"	-61, -42"
RB1	1,65806304175398E-06	"-427, -419; -372, -364"		
AR	9,10072677406948E-06	"-264, -258"		
ESR1	3,59808198545641E-05	"-243, -236; -172, -165"		"-252, -234"
EP300	0,000044207516770034	"-253, -249; -160, -156"		
E2F4	8,93560484188649E-05	"-63, -51"		
YY1	0,000137272472965688	"-112, -107"		
JUN	0,000145495268856988	"-242, -237; -171, -166"		
NFKB1	0,00018466119032956	"-328, -319"		
FOXO3	0,000213124561795786	"-427, -420; -83, -76"		
TFAP4	0,000217822979133642	"-33, -23"		
FOXM1	0,000263850910152796	"-100, -90; -97, -87"		
E2F7	0,000307318494464526		"-62, -49"	
STAT3	0,000446026823605949		"-317, -307"	
CEBPA	0,000866264009071749	"-414, -412; -366, -364; -298, -296; -243, -241; -219, -217; -127, -125; -102, -100; -86, -84; -75, -73; -61, -59; -54, -52"		
KLF5	0,000926883466516037		"-194, -185; -40, -31"	
TFDP1	0,000939306791566952	"-217, -208; -161, -152; -39, -30"		
EGR1	0,00107678487619778	"-196, -184; -43, -31"		
PPARG	0,00110813310745395		"-303, -289; -241, -227"	
HMGB1	0,00129749516746147	"-433, -427; -327, -321; -141, -135; -48, -42"		
PAX6	0,00154277794287116	"-433, -429; -172, -168"		
JUNB	0,00159487251208501	"-242, -237; -171, -166"		
TTK	0,00161234367163465	"-263, -256"		
RELA	0,00162954362946299	"-361, -351; -329, -320"		

Continúa en la siguiente página

Cuadro 7 (CDC25A) – continuación de la página anterior

Factor de transcripción	ValorP ajustado	PROMO	FIMO	MATCH
XBP1	0,00164965519297855	“-336, -328”		
ESR2	0,00209899997718643	“-171, -163”		
TAL1	0,00233275157872159	“-436, -428; -265, -257; -173, -165”		
SP1	0,00250308182017633	“-195, -186; -194, -186; -192, -187; -42, -34; -40, -35”	“-194, -184; -98, -88; -41, -31”	
STAT1	0,00284374901310602	“-52, -43”		
INSM1	0,00288459060965896	“-367, -355; -192, -180”		
IRF1	0,00360774463345304	“-52, -44”		
MAZ	0,00390682694344346	“-44, -34”		
NR2F2	0,00401293411487857		“-243, -233”	
PBX1	0,0040748957142448		“-79, -68”	
THRA	0,00442961998064695	“-242, -234; -171, -163”		
E2F5	0,0049703060427525	“-62, -55”		
PAX2	0,0050085471121516	“-245, -239; -174, -168”		
FOS	0,00553738476804533	“-242, -236; -171, -165”		
TCF4	0,00605968548271807	“-368, -362; -300, -294”		
ZEB1	0,00726692885053493	“-436, -429; -265, -258; -172, -165”		
RBPJ	0,00790805650798976		“-318, -309”	
FOXO1	0,00796145999668019	“-89, -79”		
THRB	0,00908728846647944	“-171, -162”		
LEF1	0,0104018338725288	“-369, -362; -301, -294”		
NR1H4	0,0106686143631195		“-245, -235”	
ASCL2	0,0107506993560628		“-33, -24; -30, -21”	
CEBPB	0,0121324403021432	“-414, -409; -301, -296; -86, -81”		
MYB	0,0124488352290208	“-353, -345; -350, -343; -65, -58”		“-351, -341; -66, -56”
ETV4	0,0128858707831518	“-51, -44”		
JUND	0,0154444773070046	“-242, -236; -171, -165”		
HBP1	0,016167537444871	“-308, -298”		
GABPA	0,0163824599720465	“-316, -306”		
KLF1	0,0164518968443601		“-40, -30”	
WT1	0,0166964694595916	“-404, -398”		
ETS1	0,0169670937109609	“-431, -425; -272, -267”		
NFE2L1	0,0170055447637748	“-172, -167”		

Continúa en la siguiente página

Cuadro 7 (CDC25A) – continuación de la página anterior

Factor de transcripción	ValorP ajustado	PROMO	FIMO	MATCH
POU2F1	0,0181246599872624			“-252, -231; -224, -203”
RARA	0,0185306045443622	“-243, -234; -172, -163”	“-241, -225; -136, -120”	
NR2C2	0,0191849174795626		“-241, -227”	
NFYA	0,0193774893636631	“-131, -123; -128, -121; -79, -71; -76, -69”	“-131, -121; -79, -69	-133, -114; -130, -111; -81, -62; -78, -59”
STAT5A	0,0197117401749652	“-430, -427; -327, -324; -317, -314; -138, -135; -67, -64; -48, -45”		
TCF3	0,0200265945086044	“-201, -195; -28, -22”		
FOXP3	0,0209054078166723	“-226, -218”		
STAT6	0,0218403982307975	“-317, -309”		
HNF1B	0,0240836044016713	“-11, -4”		
RXRA	0,0252068788770352	“-242, -234; -171, -163”		
ASCL1	0,0293324141468732		“-34, -22; -32, -20; -31, -19”	
SF1	0,0299205076089272	“-242, -233; -172, -162”		
ATF3	0,0301833265626682	“-242, -235; -171, -164”		
CTCF	0,0316368476692821		“-395, -377; -193, -175”	
NRF1	0,0322345808736546		“-374, -364”	
PTF1A	0,0361465513908264	“-334, -325”		
SIX3	0,0363273400822686		“-222, -206”	
NF1	0,0368771030444086	“-316, -311; -256, -251; -167, -162; -127, -123 ; -61, -57; -56, -52; -26, -22”		
PPARA	0,0376020725338466	“-242, -236; -171, -165”		
MYOG	0,0394828369386035	“-30, -24; -29, -23”	“-31, -21; -30, -20”	
ZIC2	0,0413910373518685	“-443, -441; -434, -432; -379, -377; -263, -261; -240, -238; -169, -167; -104, -102; -73, -71; -4, -2”		

Continúa en la siguiente página

Cuadro 7 (CDC25A) – continuación de la página anterior

Factor de transcripción	ValorP ajustado	PROMO	FIMO	MATCH
HNF4A	0,0414447744575829	“-301, -290; -241, -230”	“-199, -184	-372, -353; -304, -285; -295, -276; -244, -225”
TFAP2C	0,0425369835191388		“-391, -380”	
ELK1	0,0459601232253142	“-431, -427; -318, -314; -139, -135; -70, -66; -68, -64; -48, -44”		“-70, -56”
TFAP2A	0,0460733549987372	“-423, -418; -388, -380; -412, -407; -373, -368; -271, -266”	“-391, -380”	
ZBTB33	0,0469700068779998		“-161, -147; -158, -144”	
CEBPD	0,0490923696665039	“-415, -413; -244, -242; -222, -220; -128, -126; -108, -106; -76, -74; -68, -66; -7, -5”		

Cuadro 7: Factores de transcripción para el promotor CDC25A con las posiciones correspondientes al inicio de transcripción y el valor-p ajustado de ToppGene.

Factor de transcripción	ValorP ajustado	PROMO	FIMO	MATCH
TP53	7,17357284685249E-11	“-120, -114”		
FOS	9,21056051039759E-06		“-321, -306”	
NFKB1	7,31041169372837E-05	“-352, -342; -196, -186”		
JUN	0,000114297970926214		“-323, -311”	
ETS1	0,00020500225195961	“-370, -364”	“-371, -362”	“-371, -361; -61,-51”
XBP1	0,000205365468611274	“-313, -308”		
E2F1	0,000411275879295681	“-314, -307; -97, -90”		
E2F4	0,000455543332426034		“-207, -197”	

Continúa en la siguiente página

Cuadro 8 (CDC25B) – continuación de la página anterior

Factor de transcripción	ValorP ajustado	PROMO	FIMO	MATCH
RELA	0,00069855868561941		“-140, -131”	
PGR	0,00113567079419186	“-44, -38”		
GRB2	0,0014183187132315	“ -167, -163; -70, -66”		
SP1	0,00248145469611283		“-107, -97; -95, -85”	
NFATC1	0,00333475961933427	“-141, -133”		
NF1	0,0035574626312006	“-356, -349; -252, -245; -222, -215”		
HOXD10	0,00388161403787424	“-44, -35”		
TP73	0,00447945564835539		“-325, -308”	
EGR1	0,0046236647789385		“-190, -177; -134, -121”	
AHR	0,00501139173742216	“-240, -230”		
ETV6	0,00512598795710895		“-371, -362; -16, -7”	
TFDP1	0,00555293433832837		“ -97, -87”	
RXRA	0,00697117128326197	“-382, -376; -83, -77”		
ETV5	0,00746594564087411		“-371, -362”	
STAT1	0,00791976512089565	“-145, -136”		
FOXP3	0,00905424062997717	“-380, -375; -277, -272; -40, -35”		
JUNB	0,0109674147077915		“-320, -311”	
HOXD9	0,0117821270883359	“-44, -35”		
CREB1	0,012048069568065		“-321, -310”	
SP8	0,0129399254596556		“-107, -96”	
EHF	0,0130184116116444		“-371, -360”	
KLF5	0,0141860024297666		“-350, -341; -184, -175; -107, -98; -95, -86”	
CTCFL	0,0144212944986768		“-201, -188”	
JUND	0,0145757145859219		“323, -309, -322, -309; -320, -311”	
RARB	0,0150365630633315	“ -383, -374”		
GCM1	0,0152079008389547		“-114, -104”	
ASCL1	0,0159403531247492		“-50, -38; -9, 3; -8, 4”	
ZEB1	0,0160200920261033		“-326, -316”	
JDP2	0,0169000514518661		“-321, -310”	

Continúa en la siguiente página

Cuadro 8 (CDC25B) – continuación de la página anterior

Factor de transcripción	ValorP ajustado	PROMO	FIMO	MATCH
MYOG	0,0174366594632067		“-7, 3”	
KLF4	0,0176722968209573		“-186, -176”	
ZFX	0,0177530316820438		“-217, -204; -110, -97”	
POU4F2	0,0180933302786905		“-31, -16”	
RARA	0,0197695848599462	“-304, -292”	“-302, -286”	
IRF1	0,0204241898400883	“-145, -137”		
ETV1	0,0210460551616291		“-371, -362”	
TFAP2B	0,0230588809892639		“-311, -299; -287, -276; -117, -105”	
CEBPB	0,0239119782439381	“-375, -372; -276, -273; -119, -116; -1, 2”		
FOSB	0,0242116248115102		“-321,-311”	
NFATC2	0,0259636563861469	“-142, -133”		
ZIC4	0,0260737061104899		“-312, -298”	
TFAP2A	0,0293497105467638	“-294, -289; -219, -214; -145, -140; -132, -127; -124, -119”	“-311, -299; -287, -276; -117, -105”	
NKX2-5	0,029943404460143			“-168, -160”
HNF1A	0,0311252569607228	“-411, -403; -40, -32 ”		
NR1H3	0,0315208037237144		“-294, -276”	
ELF3	0,0315525032223899		“-372, -360”	
SPDEF	0,0343288769527826		“-371, -361”	
ATF3	0,0348777530312249	“-320, -313”		
NR3C1	0,0349347568093754	“-350, -346; -267, -263; -233, -229; -123, -119”		
THRB	0,0349654718097267	“-336, -328; -326, -318”		
NR2C2	0,0381137148340638		“-124, -110”	
TBP	0,0384256481665194	“-35, -29”	“-27, -13”	
POU2F2	0,041242975989803	“-410, -400”		
FOXA1	0,0447185861764583	“-226, -219; -142, -135; -38, -31”		
MEF2A	0,0484863015162203	“-38, -28 ”		

Cuadro 8: Factores de transcripción para el promotor CDC25B con las posiciones correspondientes al inicio de transcripción y el valor-p ajustado de ToppGene.

Factor de trans-cripción	ValorP ajustado	PROMO	FIMO	MATCH
TP53	1,20208287768264E-09	"-144, -138; -81, -75; -68, -62"		
NFKB1	8,33467826344414E-06	"-263, -253"		
GRB2	1,82708726673697E-05	"-29, -25"		
STAT1	2,29797162158984E-05	"-450, -441"		
FOS	0,000113925618509203		"-214,-199; -54, -39; -50, -35 "	
TFAP2A	0,000200788354229431	"-489, -484; -454, -449; -437, -432; -407, -402; -393, -388; -292, -287; -201, -196; -61, -56"	"-437, -426"	
NF1	0,000336236658390554	"-412, -405; -286, -279"		
YY1	0,000362381243301679	"-330, -327"		
AHR	0,000540337628454401	"-70, -61"		
E2F1	0,000663996583590176	"-147, -140; -99, -92; -79, -72"		
JUNB	0,000692615318351519		"-49, -40"	
ESR1	0,00111771594097909	"-324, -320"		
RORA	0,00115235539908554		"-323, -314"	
REST	0,00162769783197181		"-142, -122"	
FOXP3	0,0019115672365263	"-220, -215"		
GLI2	0,00257062915781836		"-349, -338"	
NR3C1	0,00262264249625699	"-453, -449; -436, -432; -74, -70; -36, -32"		
ETS1	0,00280453851657159	"-449, -443"		"-259, -240"
HNF4A	0,00287984644849559			"-414, -395"
LEF1	0,00308374124855948	"-4, 3"		
RXRA	0,00312086969210945	"-381, -375; -249, -243; -283, -277; -183, -177; -161, -155"		
RARA	0,00327402342249461		"-155, -139"	
CREB1	0,00413599263279951	"-50, -42; -49, -41"	"-50, -39"	"-52, -42; -50, -40; -48, -38"

Continúa en la siguiente página

Cuadro 9 (FSCN1) – continuación de la página anterior

Factor de transcripción	ValorP ajustado	PROMO	FIMO	MATCH
XBP1	0,00430880912041554	“-261, -256; -207, -202; -28, -23”		
WT1	0,00471033094206064	“-197, -189; -136, -128”		
ATF2	0,00499127372733543	“-50, -41”		“-50, -41”
IKZF1	0,00542010088494993	“-459, -447”		
AR	0,00669576833647556	“-278, -270; -210, -202; -31, -23”		
JUN	0,00719223751029363		“-52, -40; -49, -37”	
ATF3	0,00748281711350929	“-324, -317; -49, -42”	“-46, -39”	
NFATC2	0,00751593835630859	“-453, -445”		
KLF4	0,00757981777161643		“-70, -60”	
NFATC1	0,00764734479152684	“-454, -445”		
EBF1	0,00922487280619766	“-88, -78”		
IRF1	0,00928941266945926	“-449, -441”		
ARNT	0,00937314134448275	“-70, -61”		
CTCF	0,00992136697058987		“-403, -385”	
KLF5	0,0107680993619276		“-81, -72; -68, -59”	
SP8	0,0110765963644632		“-70, -59”	
E2F4	0,0112974278259523		“-234, -224; -99, -89”	
PAX5	0,0113501895773486		“-67, -56”	
PPARA	0,0114125403351324		“-324, -307”	
SP1	0,0176199656423104		“-376, -366; -269, -259; -82, -72”	
STAT4	0,017869292518445	“-450, -445”		
SREBF2	0,0184110795890541		“-163, -154”	
RREB1	0,0186505647111622		“-254, -235; -253, -234”	
CEBPB	0,0188226640582155	“-408, -405; -319, -316; -302, -299; -286, -283; -286, -283; -219, -216; -2, 1”		
TCF4	0,0216031091896111	“-4, 2”		
JUND	0,0225031422312965		“-50, -36”	
SP2	0,0236231134030711		“-376, -362; -86, -72”	
RXRG	0,0239813209778472		“-323, -310”	

Continúa en la siguiente página

Cuadro 9 (FSCN1) – continuación de la página anterior

Factor de transcripción	ValorP ajustado	PROMO	FIMO	MATCH
NR4A2	0,0239974257200334		“-405, -398; -324, -317”	
ASCL1	0,0259343429566832		“-141, -129”	
FOSL2	0,0291868634169145		“-52,-38; -51, -37; -50, -40; -49, -39”	
THRA	0,0304790554793035			“-328, -313”
THRB	0,0313220503074284	“-215, -207; -164, -156; -44, -36”		
CR1	0,0319883297170478			“-359, -349”
MEIS2	0,0323288521151817		“-467, -460”	
NR2C2	0,032638619315264		“-386, -372; -323, -309”	
ETV5	0,0331739501305099		“-259, -250”	
JDP2	0,0335479178381549		“-50, -39”	
ETV4	0,0350404107131816	“-261, -253”		
HOXB13	0,0358438096277251		“-35, -36”	
TFAP2B	0,0361375182965071		“-437, -426”	
TBP	0,0389593275516582	“-38, -29; -34, -28”	“-34, -20”	
CTCFL	0,0400604258357119		“-202, -189”	
RARB	0,0416573500285079	“-250, -241; -223, -214”		
EGR3	0,04191422191594	“-379, -367”	“-74, -60”	
SREBF1	0,0429622450550528		“-163, -154”	
HOXA13	0,0444090855665714		“-35, -26”	
RXRB	0,0462013945695109		“-323, -310”	
TFDP1	0,0469945374835014		“-234, -224; -99, -89; -80, -70”	

Cuadro 9: Factores de transcripción para el promotor FSCN1 con las posiciones correspondientes al inicio de transcripción y el valor-p ajustado de ToppGene.

Factor de trans-cripción	ValorP ajustado	PROMO	FIMO	MATCH
TP53	2,13828228166069E-05	"-303, -297; -46, -40; -30, -24; -13, -7"		
AR	5,44330229713674E-05	"-179, -173"		
E2F1	0,000114541207416052	"-302, -296; -195, -189; -137, -131"		
RARB	0,00020161009275288	"-368, -357"		
LEF1	0,00041015649801901	"-372, -365"		
STAT1	0,000435494173668394		"-80, -66"	
ETS1	0,000571137404939637	"-207, -198; -192, -187"		"-206, -185"
NFKB1	0,000596547922880664	"-259, -248; -134, -125"		
RXRA	0,000657070295259632	"-368, -357; -29, -19"		
NF1	0,000911384309951302	"-365, -361; -118, -113; -117, -113; -15, -8; -4, 0"		
NFATC1	0,000942913945405355	"-162, -154; -81, -73"		
CEBPA	0,000952190238138861	"-235, -230; -234, -229; -233, -228; -231, -229; -208, -206; -208, -203"	"-235, -224"	"-235, -222"
PAX8	0,000998298469045311	"-332, -327"		
PPARA	0,00120770031196149	"-29, -19"		
PAX6	0,00134427948800919	"-273, -269; -243, -239"		
PAX5	0,00144207135607333	"-316, -310"		
EP300	0,00148473913478109	"-350, -346; -339, -335"		
RELA	0,00159724738204836	"-135, -126"		
ASCL1	0,00170022482309129		"-117, -105; -115, -103"	
NKX2-1	0,0019457904843706	"-270, -264; -184, -178"		
STAT5A	0,00210904705377946	"-204, -201; -162, -159; -133, -130; -81, -78"		
ARNT	0,00240295674382596	"-320, -312"		
POU3F2	0,00251707323202965	"-351, -345"		
VDR	0,00256645306572611	"-275, -272; -245, -242; -236, -233; -218, -215; -196, -193; -55, -52"		
KLF4	0,0030209207817703	"-296, -284"		
HES1	0,00302980314029822	"-318, -312"		
NFE2L2	0,00304163066204832	"-181, -175"		
AHR	0,0032355454999059	"-319, -311; -316, -309"		

Continúa en la siguiente página

Cuadro 10 (GLI1) – continuación de la página anterior

Factor de transcripción	ValorP ajustado	PROMO	FIMO	MATCH
MSX1	0,00333247728498032	“-327, -325; -314, -312; -234, -232; -219, -217; -147, -145; -53, -51”		
MECOM	0,00334057792500719			“-104, -83; -80, -59”
TFAP2A	0,00342292440247338	“-376, -371; -191, -186”	“-37, -25”	
POU1F1	0,00368285436361782	“ -337, -333”		
SP1	0,00369149004933367	“-304, -295; -303, -298; -196, -186”	“-304, -294; -265, -255; -195, -185; -168, -158”	
RARG	0,00374070756660105	“-331, -324; -148, -141”		
LHX1	0,00407685764895493	“-370, -367; -347, -344; -325, -322”		
CEBPB	0,00410239093115816	“ -234, -229; -92, -85; -90, -83”		
PGR	0,00436346319396286	“-315, -309; -313, -307”		
FOXO3	0,00437762434448241	“ -290, -283; -232, -225”		
GLIS2	0,00439398521504986		“-304, -291; -253, -240; -60, -47”	
EGR1	0,00446680262291854	“-306, -294”		
FOXM1	0,0048290220306153	“ -339, -333; -373, -368”		
DDIT3	0,00502039689169576		“-234,-225”	
WT1	0,00509789157787055	“ -45, -37; -34, -26 ”		
MYOG	0,00511595636336404			“-15, 14”
E2F4	0,00565097186434904		“-302, -292; -137, -127; -20, -10”	
TCF4	0,0070178217743263	“-372, -366”		
TCF3	0,0072764841821874	“ -356, -350; -182, -176”		
ETV4	0,00732942577219997	“-190, -183”		
KLF5	0,00786509476053221		“-303, -294; -264, -255; -167, -158”	
MYOD1	0,00790724880894411	“ -321, -314”		
TAL1	0,00826646534303088	“-2, 6”		
PAX9	0,0112517526011865	“ -219, -213”		
INSM1	0,0116688569854118		“-242, -231”	
HOXD10	0,013182220368785	“-159, -150”		

Continúa en la siguiente página

Cuadro 10 (GLI1) – continuación de la página anterior

Factor de transcripción	ValorP ajustado	PROMO	FIMO	MATCH
ZIC3	0,015061810380807		“-23, -9”	
POU3F1	0,0151177046676514	“-163, -157”		
NFATC4	0,0159510564885061	“ -81, -72”		
EWSR1	0,0178298246644708		“-173, -156”	
SP3	0,0183607915620701	“ -304, -296”	“-304, -294; -298, -288; -297, -287”	
POU2F1	0,0199399165079154	“ -92, -83”		
TEAD2	0,0239801683376659	“-304, -296”		
EGR3	0,0264035158998411		“-309, -295; -303, -289; -301, -287; -297, -283”	
FOXD3	0,0264136234040354			“-78, -65”
MAFB	0,0269363911895151		“-287, -76”	
TTK	0,028674963375187	“-244, -237”		
SP8	0,0292186198049805		“-300, -289; -298, -287”	
MYB	0,0294439661115906	“-180, -173”		“-179, -169”
PLAG1	0,0297085026379084		“-303, -290; -274, -261; -244, -231”	
TFAP2B	0,0303267551908553	“ -250, -239; -241, -230”	“-271, -260; -37, -25”	
TFDP1	0,0306772232394597	“-137, -128; -20, -11”	“-137, -127; -20, -10”	
NR2C2	0,032213301879094		“-305, -291; -300, -286; -150, -136”	
KLF6	0,0326246158199474	“ -250, -240; -241, -231”		
CDX1	0,032848365614299	“-336, -333; -286, -283; -228, -225”		
IRF1	0,0328957206638373	“-85, -77”	“-84, -64”	
ZNF423	0,0335105693569441		“-376, -362”	
MZF1	0,0392304991230143		“-23, -9”	
ZIC4	0,0399339114638826		“-262, -253”	
TFAP2C	0,0410791004612802		“-271, -260; -37, -25”	

Continúa en la siguiente página

Cuadro 10 (GLI1) – continuación de la página anterior

Factor de transcripción	ValorP ajustado	PROMO	FIMO	MATCH
CEBPD	0,0420559963685923	“-350, -348; -337, -335; -230, -228; -159, -157”		
HOXD9	0,0422781539415785	“-159, -150”		
NKX6-2	0,0437341094852443	“-163, -156”		
E2F6	0,0464246256321999		“-170, -160; -166, -156; -137, -127; -20, -10”	
NFATC3	0,0474730498157291	“-81, -70”		
HOXA3	0,0485460194972342	“-323, -319”		

Cuadro 10: Factores de transcripción para el promotor GLI1 con las posiciones correspondientes al inicio de transcripción y el valor-p ajustado de ToppGene.

Factor de transcripción	ValorP ajustado	PROMO	FIMO	MATCH
TP53	5,85895317817275E-08	“-375, -369; -372, -366”		
NFE2L2	4,17099956029343E-06		“-462, -448”	
PPARG	0,000036605851631788		“-258, -244; -180, -166”	
FOS	0,000167599293366849		“-459, -449”	
JUN	0,000380377619455086	“-458, -452”	“-463, -450”	
ESR2	0,000442071054509108		“-309, -295”	
NR3C1	0,000579800492858551	“-690, -686; -532, -528; -378, -374; -323, -319; -266, -260; -256, -252; -167, -163; -38, -34”		
XBP1	0,000632275731883847	“-591, -586; -375, -370; -371, -366”		
AR	0,000782141303374173	“-63, -55; -40, -32”		
PGR	0,00103221304694179	“-645, -639; -573, -567; -523, -517”		
IRF1	0,00207208118992874	“-652, -644”		
CEBPA	0,00227939489423423	“-639, -633; -237, -231”		

Continúa en la siguiente página

Cuadro 11 (NQO1) – continuación de la página anterior

Factor de transcripción	ValorP ajustado	PROMO	FIMO	MATCH
FOXA1	0,00241570644570821	“-523, -516; -477, -470; -267, -260”		
KLF4	0,00249638818317721		“-245, -235”	
HNF4A	0,00269415262630779			“-385, -369”
NR1H2	0,00296801994175577		“-184, -168”	
BACH1	0,00389865195492734		“-461, -447; -123, -109”	
E2F1	0,0046648159638113	“-181, -174; -110, -103; -77, -70”		
JUND	0,00481907114122815		“-459, -449”	
STAT1	0,00484115461330659		“-133, -123”	
LEF1	0,00528540985263415	“-536, -529”		
HIC1	0,00681570878377347		“-370, -362”	
RARG	0,00713518044835315		“-384, -369”	
EWSR1	0,00835596534351324		“-146, -129”	
THRB	0,00843048294862037	“-182, -174”		
NEUROD1	0,00933171249368769		“-299, -287”	
NF1	0,00940770613168418	“-391, -384; -346, -339”		
FOSL2	0,00988441454084565		“-737, -727; -460, -450”	
TFAP2A	0,0109788615389371	“-726, -721; -440, -435; -286, -281; -147, -142; -87, -82”	“-440, -430; -315, -304; -278, -267; -150, -140”	
SP1	0,011183640025212		“-231, -221; -162, -152; -141, -131; -108, -98; -75, -65”	
RARA	0,0115293488979964		“-384, -367”	
TFAP2B	0,0120097998967437		“-440, -430; -315, -304; -150, -140”	
MAF	0,0121450497374005		“-464, -444; -461, -447; -458, -444; -126, -106; ”	“-463, -441”

Continúa en la siguiente página

Cuadro 11 (NQ01) – continuación de la página anterior

Factor de transcripción	ValorP ajustado	PROMO	FIMO	MATCH
CEBPB	0,0123492434189149	“ -663, -660; -644, -641; -568, -565; -536, -533; -534, -531; -522, -519; -319, -316; -288, -285; -236, -233; -189, -186; -67, -64”		
FOSL1	0,0134457685162285		“-459, -449”	“-458, -449”
FOSB	0,0137872765459522		“-459, -450; -458, -449”	“-458, -449”
MYB	0,0143001772376347	“-689, -679; -194, -187”		“-719, -709; -687, -677; -50, -40”
RXRA	0,0173713308967988	“-586, -580; -179, -173”		
PAX5	0,0194033196284552	“-375, -369; -372, -366; -292, -286; -223, -217; -74, -68; -45, -39”	“ -78, -67”	
KLF9	0,0199916503275922		“-251, -239; -249, -237; -247, -235”	
ETS1	0,0211455918697501	“ -531, -525; -134, -128”	“ -418, -409”	“-418, -402”
ZBTB7A	0,0213863978829902		“-420, -408; -137, -125”	
JUNB	0,0214121013334736		“-460, -450; -458, -448”	
NR2C2	0,0225633670284666		“ -180, -166”	
RARB	0,0238140889894959		“-385, -370”	
BACH2	0,0238895986872945		“-459, -446”	
STAT6	0,0251521920213095		“-694, -680”	
MSX1	0,0272031910954921		“ -637, -630”	
HES1	0,0286098108054962		“-3, 6”	
NKX2-5	0,0302873542182932		“-170, -160”	“-365, -358”
GRB2	0,0354102785557281	“ -637, -633; -625, -621; -520, -516; -397, -393”		
MECOM	0,0359481033956208			“-30, -20”
MTF1	0,0379020695996408		“-306, -293”	
MAZ	0,0380569845089093	“ -232, -220; -142, -130”		
STAT4	0,0399493160678426	“-648, -643; -632, -627”		

Continúa en la siguiente página

Cuadro 11 (NQO1) – continuación de la página anterior

Factor de transcripción	ValorP ajustado	PROMO	FIMO	MATCH
BSX	0,0402948319081318		“-637, -630; -625, -618”	
DMBX1	0,0402948319081318		“-428, -412”	
EBF1	0,0414625648825568	“-584, -574; -315, -305”	“-351, -338; -220, -207”	
OTX1	0,0422947962116402		“-424, -417”	
GCFC2	0,0454891673722491	“-733, -725; -300, -292 ”		
YY1	0,0457258959541861	“-651, -648; -625, -622; -564, -561; -397, -394; -294, -291”		
ARX	0,0464760516190524		“-641, -625”	
EN1	0,0479241867102774		“-637, -630”	
TBP	0,048597524836193	“-265, -259”		

Cuadro 11: Factores de transcripción para el promotor NQO1 con las posiciones correspondientes al inicio de transcripción y el valor-p ajustado de ToppGene.

Factor de transcripción	PROMO	FIMO	MATCH
TP53	“-177, -171; -33, -27”	“-314, -297”	
POU2F1	“-289, -279”	“-288, -277”	“-289, -274; -288, -273;-287, -272; -89, -74; -68, -53”
FOXA1	“-298, -291; -285, -278; -193, -186; -186, -179; -79, -72; -52, -45”	“-60, -46”	
TBP	“-300, -294; -191, -185; -188, -182; -73, -67; -48, -42”	“-36, -22”	
MYB	“-305, -298; -125, -118”	“-126, -117”	
FOXD3		“-58, -47”	“-58, -43”
NKX2-5		“-248, -238”	“-281, -266”
FOS	“-270, -262; -264, -256; -62, -54; -23, -15; -20, -11”	“-13, 2”	“-22, -7”
STAT4	“-284, -279”		

Continúa en la siguiente página

Cuadro 12 (VWA2) – continuación de la página anterior

Factor de transcripción	PROMO	FIMO	MATCH
SOX5			“-119, -104”
SOX9		“-118, -110”	“-121, -106”
XBP1	“-266, -261;-81, -76”		
MNX1		“-101, -92”	
HNF4A	“-61, -49”		“-150, -135”

Cuadro 12: Factores de transcripción para el promotor VWA2 con las posiciones correspondientes al inicio de transcripción.

12.7. Resultado de la clasificación de los factores de transcripción según las variables definidas en el estudio

A continuación, se presenta la tabla de todos los factores encontrados en los ocho promotores ordenados alfabéticamente, categorizados en función de las variables definidas en el estudio. La tabla 13 representa los resultados correspondientes. Por otro lado, la tabla 14 describe a los factores encontrados en función de su rol en la enfermedad CCR. Esto es importante ya que como bien se mencionó anteriormente, varios factores no cumplen con las variables definidas, mas su rol es fundamental en el desarrollo de la enfermedad ya que actúa como subunidad, por ejemplo. Esta tabla se hizo con el fin de que los investigadores puedan rápidamente comprender el por qué de la presencia de dicho factor en la posición en la cual se encontró.

TF	EDC	EDT	AV	ARI	TF	EDC	EDT	AV	ARI
AHR	X	X	!X	!X	AR	X	!X	!X	!X
ARNT	X	!X	!X	!X	ARX	!X	!X	!X	!X
ASCL1	!X	!X	!X	!X	ASCL2	!X	!X	!X	!X
ATF1	X	!X	!X	!X	ATF2	!X	!X	!X	!X
ATF3	!X	!X	!X	!X	BACH1	X	!X	!X	!X
BACH2	!X	!X	!X	X	BSX	!X	!X	!X	!X
CDX1	X	X	!X	!X	CDX2	X	X	!X	!X
CEBPA	!X	!X	!X	X	CEBPB	!X	!X	!X	X
CEBPD	!X	!X	!X	X	CR1	!X	!X	!X	X
CREB1	X	!X	!X	!X	CTCF	!X	!X	!X	!X
CTCFL	X	!X	!X	!X	CUX1	X	!X	!X	!X
DDIT3	X	!X	!X	X	DMBX1	!X	!X	!X	!X
DSTN	!X	!X	!X	!X	E2F1	X	!X	X	!X

Continúa en la siguiente página

Cuadro 13 – continuación de la página anterior

TF	EDC	EDT	AV	ARI	TF	EDC	EDT	AV	ARI
E2F4	X	!X	!X	!X	E2F5	X	!X	!X	!X
E2F6	!X	!X	!X	!X	E2F7	!X	!X	!X	!X
EBF1	!X	!X	!X	!X	EGR1	X	!X	!X	!X
EGR3	!X	!X	!X	X	EHF	X	!X	!X	!X
ELF3	X	X	!X	!X	ELK1	X	!X	!X	!X
EN1	!X	!X	!X	!X	EP300	X	!X	X	!X
ESR1	X	!X	!X	!X	ESR2	X	!X	!X	!X
ETS1	X	!X	!X	!X	ETV1	X	!X	!X	!X
ETV4	X	!X	X	!X	ETV5	X	!X	!X	!X
ETV6	!X	!X	!X	!X	EZH2	X	!X	X	X
EWSR1	X	!X	!X	!X	FLI1	X	!X	!X	!X
FOS	X	!X	!X	!X	FOSB	!X	!X	!X	!X
FOSL1	!X	!X	!X	!X	FOSL2	!X	!X	!X	!X
FOXA1	X	!X	!X	!X	FOXD3	X	!X	!X	!X
FOXM1	X	!X	!X	!X	FOXO1	X	!X	!X	!X
FOXO3	X	!X	!X	!X	FOXP3	!X	!X	!X	X
GABPA	!X	!X	!X	!X	GCFC2	!X	!X	!X	!X
GCM1	!X	!X	!X	!X	GLI2	!X	!X	!X	!X
GLIS2	!X	!X	!X	!X	GRB2	X	!X	X	!X
HBP1	!X	!X	!X	!X	HES1	!X	!X	!X	!X
HIC1	X	!X	!X	!X	HMGB1	!X	!X	!X	X
HNF1A	X	X	!X	!X	HNF1B	!X	!X	!X	!X
HNF4A	X	X	!X	!X	HOXA3	!X	!X	!X	!X
HOXA13	X	!X	!X	!X	HOXB13	!X	X	!X	!X
HOXD10	!X	!X	!X	!X	HOXD9	!X	!X	!X	!X
IKZF1	X	!X	!X	!X	INSM1	X	!X	!X	!X
IRF1	X	!X	!X	X	ITGAM	X	!X	!X	X
JDP2	!X	!X	!X	!X	c-JUN	X	!X	!X	!X
JUNB	!X	!X	!X	!X	JUND	!X	!X	!X	!X
KLF1	!X	!X	!X	!X	KLF4	X	!X	!X	!X
KLF5	X	X	!X	!X	KLF6	X	!X	!X	!X
KLF9	!X	!X	!X	!X	LEF1	X	!X	!X	X
LHX1	X	!X	!X	!X	MAF	X	!X	!X	X
MAFB	X	!X	!X	!X	MAZ	!X	!X	!X	!X
MECOM	X	!X	!X	!X	MEF2A	X	!X	!X	!X
MEIS2	!X	!X	!X	!X	MNX1	X	!X	!X	!X
MSX1	!X	!X	!X	!X	MTF1	!X	!X	!X	!X
MTUS1	!X	!X	!X	!X	MYB	X	!X	!X	!X
MYC	X	!X	!X	!X	MYOD1	X	!X	!X	!X
MYOG	!X	!X	!X	!X	MZF1	!X	!X	!X	!X
NEUROD1	!X	!X	!X	!X	NF1	X	!X	!X	!X
NFATC1	!X	!X	!X	!X	NFATC2	X	!X	!X	!X
NFATC3	!X	!X	!X	!X	NFATC4	!X	!X	!X	X

Continúa en la siguiente página

Cuadro 13 – continuación de la página anterior

TF	EDC	EDT	AV	ARI	TF	EDC	EDT	AV	ARI
NFE2L1	!X	!X	!X	!X	NFE2L2	X	!X	!X	!X
NFKB1	X	!X	!X	!X	NFYA	!X	!X	!X	!X
NKX2-1	!X	!X	!X	!X	NKX2-5	!X	!X	!X	!X
NKX6-2	!X	!X	!X !X	NR1H2	X	!X	!X	!X	
NR1H3	!X	!X	!X	!X	NR1H4	!X	!X	!X	X
NR1I2	X	X	!X	X	NR2C2	!X	!X	!X	!X
NR2F2	!X	!X	!X	!X	NR3C1	!X	!X	!X	!X
NR4A2	X	!X	!X	!X	NRF1	!X	!X	!X	!X
OTX1	!X	!X	!X	!X	PAX2	!X	!X	!X	!X
PAX5	X	!X	!X	!X	PAX6	!X	!X	!X	!X
PAX8	X	!X	!X	!X	PAX9	!X	!X	!X	!X
PBX1	X	!X	!X	!X	PGR	X	!X	!X	!X
PLAG1	X	!X	!X	!X	POU1F1	!X	!X	X	!X
POU2F1	!X	!X	X	!X	POU2F2	X	X	!X	!X
POU3F1	!X	!X	!X	!X	POU3F2	!X	!X	!X	!X
POU4F2	!X	!X	!X	!X	PPARA	!X	!X	!X	X
PPARG	X	!X	!X	!X	PTF1A	!X	!X	!X	!X
RARA	X	!X	!X	!X	RARB	!X	!X	!X	!X
RARG	X	!X	!X	!X	RB1	X	!X	X	!X
RBPJ	!X	!X	!X	!X	RELA	X	!X	X	!X
REST	!X	!X	!X	!X	RORA	!X	!X	!X	!X
RREB1	X	!X	!X	!X	RXRA	!X	!X	X	X
RXRB	!X	!X	!X	!X	RXRG	!X	!X	!X	!X
SIX3	!X	!X	!X	!X	SMAD1	X	!X	!X	!X
SNAI1	!X	!X	!X	!X	SOX2	X	!X	!X	!X
SOX5	X	!X	!X	!X	SOX9	X	!X	!X	!X
SP1	X	!X	!X	!X	SP2	!X	!X	!X	!X
SP3	!X	!X	!X	!X	SP8	!X	!X	!X	!X
SPDEF	!X	!X	!X	!X	SREBF1	!X	!X	!X	!X
SREBF2	!X	!X	!X	!X	STAT1	X	!X	!X	!X
STAT3	X	!X	!X	!X	STAT4	X	!X	!X	!X
STAT5A	X	!X	!X	!X	STAT6	X	!X	!X	X
TAL1	X	!X	!X	X	TBP	!X	!X	!X	!X
TCF3	X	!X	!X	!X	TCF4	!X	!X	!X	!X
TEAD2	X	!X	!X	!X	TFAP2A	!X	!X	!X	!X
TFAP2B	!X	!X	!X	!X	TFAP2C	!X	!X	!X	!X
TFAP4	!X	!X	!X	!X	THRA	!X	!X	!X	!X
THRB	!X	!X	!X	!X	TP53	X	!X	!X	!X
TP73	X	!X	!X	!X	TTK	!X	!X	!X	!X
VDR	X	X	!X	!X	WT1	X	!X	!X	!X
XBP1	X	!X	!X	!X	YY1	X	!X	X	!X
ZBTB33	!X	!X	!X	!X	ZBTB7A	!X	!X	!X	!X
ZEB1	!X	!X	!X	!X	ZFX	!X	!X	!X	X

Continúa en la siguiente página

Cuadro 13 – continuación de la página anterior

TF	EDC	EDT	AV	ARI	TF	EDC	EDT	AV	ARI
ZIC2	!X	!X	!X	!X	ZNF423	!X	!X	!X	!X

Cuadro 13: Resultado de todos los factores de transcripción identificados como estadísticamente significativos luego del ranqueo genético. DSTN y ADF-1 son distintas nomenclaturas para el mismo TF. **TF**: factor de transcripción; **EDC**: Específico de cáncer; **EDT**: Específico de tejido; **AV**: Actividad viral; **ARI**: Actividad de respuesta inmunológica.

Factor de transcripción	Descripción
AHR	Sobre-expresión de tejido (placenta). Proceso biológico: ritmos biológicos, ciclo celular, transcripción, regulación de la transcripción (Uniprot). Activador transcripcional activado por ligando. Se une a la región promotora de genes XRE que activa (Laura MacPherson et al., 2013)(Alvin Gomez et al., 2018). El factor AhR está altamente conservado en la secuencia FSCN1. Los factores AhR y CREB son determinantes importantes de la actividad transcripcional de la región FSCN1 -219 / -43, y que la asociación de estos factores de transcripción con la región promotora FSCN1 correspondiente está específicamente elevada en la fascia. células positivas de carcinoma de colon humano (Hashimoto et al; 2009).
AR	Sobre-expresión de tejido (hígado). El gen del receptor de andrógenos tiene más de 90 kb de longitud y codifica una proteína que tiene 3 dominios funcionales principales: el dominio N-terminal, ADN-bindi CCR tiende a ocurrir con mayor frecuencia en hombres que en mujeres, y la evidencia muestra que el CCR es un tumor relacionado con hormonas. Estos hallazgos indican que la metilación del gen del receptor de andrógenos (AR) podría ser importante para la regulación del riesgo de CCR en los diferentes sexos (Xia T et al; 2019)
ARNT	Requerido para la actividad del receptor Ah (dioxina). Esta proteína es necesaria para que la subunidad de unión al ligando se transloque del citosol al núcleo después de la unión del ligando. El heterodímero se une a la secuencia de ADN central 5'-TACGTG-3 'dentro del elemento de respuesta a la hipoxia (HRE) de los promotores de genes diana y funciona como un regulador transcripcional de la respuesta adaptativa a la hipoxia. El heterodímero ARNT:AHR se une a la secuencia de ADN central 5'-TGC GTG-3 'dentro del elemento de respuesta a dioxina (DRE) de los promotores de genes diana y activa su transcripción (Seung-Hyeon Seok et al;2017)
Continúa en la siguiente página	

Cuadro 14 – continuación de la página anterior

Factor de transcripción	Descripción
ARX	Sobre-expresión de tejido (cerebro, ovario, músculo esquelético). Proceso biológico: diferenciación, neurogénesis, transcripción, regulación de la transcripción. TF requerido para el desarrollo normal del cerebro. Puede ser importante para el mantenimiento de subtipos neuronales específicos en la corteza cerebral y la guía axonal en la placa del piso (THPA)
ASCL1	Sobre-expresión de tejido(cerebro, hígado, glándula pituitaria). Proceso biológico: diferenciación, neurogénesis, transcripción, regulación de la transcripción. TF que juega un papel clave en la diferenciación neuronal: actúa como un TF pionero, accediendo a la cromatina cerrada para permitir que otros factores se unan y activen las vías neurales (THPA)
ASCL2	Sobre-expresión de tejido(conducto deferente, placenta). Proceso biológico: diferenciación, neurogénesis. Participa en la determinación de los precursores neuronales en el sistema nervioso periférico y el sistema nervioso central (THPA)
ATF1	Esta proteína se une al elemento de respuesta AMPc (CRE), una secuencia presente en muchos promotores virales y celulares. Se une al elemento que responde a las secuencias (TRE) de HTLV-I. Media la estimulación inducida por PKA de genes informadores CRE. Reprime la expresión de FTH1 y otros genes de desintoxicación antioxidantes. ATF1 funciona como un oncogén que facilita la proliferación celular; ATF1 tiene el efecto más significativo de los genes identificados y promueve el crecimiento de xenoinjerto de CCR al afectar la apoptosis celular. ATF1 es un factor clave en el CCR modulado por una interacción promotor-potenciador (Jianbo Tian et al; 2019)
ATF2	Activador transcripcional que regula la transcripción de genes implicados en la antiapoptosis, el crecimiento celular y la respuesta al daño del ADN. Dependiendo de su compañero de unión, se une a CRE (elemento de respuesta de AMPc) o a AP-1. En el núcleo, contribuye a la transcripción global y la respuesta al daño del ADN, además de actividades transcripcionales específicas que están relacionadas con el desarrollo celular, la proliferación y la muerte (THPA). Interactúa con E1A y coloca su dominio de transactivación en un promotor viral; activador, independiente de cAMP, activador más fuerte como heterodímero con c-Jun que como homodímero (Matsuda S et al; 1991)
ATF3	Sobre-expresión de tejido(vesícula seminal). Miembro de la familia de factores de transcripción de la proteína de unión al elemento sensible / cAMP (ATF / CREB). ATF3 reprime en lugar de activar la transcripción de promotores con sitios ATF; regula la expresión de genes diana específicos (Chen BP et al; 1994)

Continúa en la siguiente página

Cuadro 14 – continuación de la página anterior

Factor de transcripción	Descripción
BACH1	Regulador transcripcional que actúa como represor o activador, según el contexto. Se une a los sitios de unión al ADN de NF-E2. Juega papeles importantes en la coordinación de la activación de la transcripción y la represión por parte de MAFK. Nrf2 y Bach1 son factores transcripcionales importantes que protegen contra las especies reactivas de oxígeno, y la aberración de estas moléculas se asoció con la transformación y progresión malignas. Bach1 fue mayor en la mucosa normal; menos en adenoma luego aumentó en carcinoma y se asoció con características de invasividad tumoral y metástasis (Heba E M El-Deek et al; 2019). BACH1 sirve como un oncogén para promover la progresión del CCR y un factor pronóstico independiente para la supervivencia y la metástasis (Guo-Dong Zhu et al; 2018)
BACH2	Sobre-expresión de tejido (tejido linfoide). Regulador transcripcional que actúa como represor o activador. Se une a los elementos de reconocimiento de Maf (MARE). Desempeña un papel importante en la coordinación de la activación de la transcripción y la represión por parte de MAFK. Induce la apoptosis en respuesta al estrés oxidativo a través de la represión del factor antiapoptótico HMOX1 (Yoshida C et al; 2007). Regula positivamente la importación nuclear de actina. Es un regulador clave de la inmunidad adaptativa, crucial para el mantenimiento de la función reguladora de las células T y la maduración de las células B (Afzali B et al; 2017)
BSX	Sobre-expresión de tejido (cerebro, glándula pituitaria, testículo). El homobox-Bsx específico del cerebro podría considerarse un marcador molecular importante para las primeras etapas embrionarias del desarrollo de la epífisis, expresándose específicamente en esta estructura neural (Mattia Cremona et al; 2004)
CDX1	La proteína de unión al ADN codificada regula la expresión génica específica del intestino y la diferenciación de enterocitos. Sobreexpresado solo en CCR según Gepia.
CDX2	La proteína codificada es un regulador principal de genes específicos del intestino involucrados en el crecimiento celular y la diferenciación. La expresión aberrante de este gen se asocia con inflamación intestinal y tumorigénesis (THPA). Sobreexpresado solo en CCR (Ha Kim et. Al. 2005) y corroborado por Gepia.
CEBPA	Sobre-expresión de tejido(mama, hígado). Activador; inhibe la proliferación celular al aumentar el nivel de p21 (WAF-1), que es un inhibidor de la quinasa dependiente de ciclina (CDK) (Harris TE et al; 2001). El TF CEBPA es crucial para la diferenciación de los granulocitos. La expresión condicional de CEBPA desencadena la diferenciación neutrofílica y no granulocitos maduros (Pabst T et al; 2001). Modula la diferenciación de macrófagos, que tienen un papel clave en la inmunidad (Yi Le et al; 2018)

Continúa en la siguiente página

Cuadro 14 – continuación de la página anterior

Factor de transcripción	Descripción
CEBPB	Sobre-expresión de tejido (músculo esquelético). Los factores CCAAT son proteínas de unión potenciadoras (o CEBPB) y están involucradas en diferentes respuestas celulares, como en el control de la proliferación celular, crecimiento y diferenciación, en el metabolismo y en la inmunidad (Martyn GE et al; 2017). Actividad de esta proteína puede modular la expresión de genes involucrados en la regulación del ciclo celular (THPA). Importante factor de transcripción que regula la expresión de genes involucrados en respuestas inmunes e inflamatorias (S Kinoshita et al; 19992) (R Chinery et al; 1997)
CEBPD	Activador de la transcripción que reconoce dos factores de ADN diferentes: la homología CCAAT común a muchos promotores y la homología central mejorada común a muchos potenciadores (Ju-Ming Wang et al; 2006). TF importante que regula la expresión de genes implicados en respuestas inmunes e inflamatorias (S Kinoshita et al; 1992). Activador transcripcional que mejora la transcripción de IL6 solo y como heterodímero con CEBPB.
CR1	Sobre-expresión de tejido (sangre, tejido linfoide). Proceso biológico: antígeno del grupo sanguíneo, receptor de la célula huésped para la entrada del virus, receptor (Uniprot). Grupo enriquecido (basófilos, eosinófilos, neutrófilos), receptor de adhesión inmunitaria de membrana que juega un papel crítico en la captura y eliminación de patógenos opsonizados por el complemento por los eritrocitos y monocitos / macrófagos (J A Schifferli et al; 1988). También desempeña un papel en la regulación inmune al contribuir, tras la unión del ligando, a la generación de células T reguladoras a partir de células T auxiliares activadas (Katalin Török etl al; 2015)
CREB1	Un regulador importante de la expresión de MHC de clase II, conocido como X2BP, coopera con RFX y NF-Y (Hasegawa SL y otros, 1991; Louis-Plence P y otros, 1991). La regulación por arriba de FSCN1 en carcinomas humanos agresivos parece tener una base multifactorial: CREB y AhR como reguladores específicos principales (Yosuke Hashimoto et al; 2009) proteína de unión al elemento de respuesta AMPc (CREB) y el receptor de hidrocarburo de arilo (AhR) en la regulación de la transcripción de FSCN1 en células de carcinoma humano, y son factores altamente conservados (Yosuke Hashimoto et al; 2009). Debido a su posición central aguas abajo de muchas vías de señalización de crecimiento, CREB tiene la capacidad de influir en la supervivencia celular, el crecimiento y la diferenciación normal, pero también de células tumorales que sugieren un potencial oncogénico de CREB. CREB podría no solo servir como marcador pronóstico, sino también como un objetivo terapéutico para los cánceres asociados con una mayor actividad de las vías de transducción de señales (André Steven et al; 2016)
Continúa en la siguiente página	

Cuadro 14 – continuación de la página anterior

Factor de transcripción	Descripción
CTCF	Se une a sitios específicos de secuencia de ADN. Participa en la regulación transcripcional uniéndose a aisladores de cromatina y evitando la interacción entre el promotor y los potenciadores y silenciadores cercanos. Actúa como un represor transcripcional que se une a los promotores del gen MYC vertebrado y el gen BAG1. La miogénesis está regulada por MyoD y otros factores de transcripción específicos del músculo que promueven un programa de expresión génica específica del músculo. CTCF interactúa físicamente con MyoD, lo que sugiere un papel para CTCF en la activación transcripcional de genes específicos del músculo (Rodrigo G. Arzate-Mejía et al; 2018).
CTCFL	Sobre-expresión en tejido (testículos). CTCFL / BORIS ha recibido atención adicional como objetivo expresado en CSC. Estas células impulsan la recurrencia del crecimiento tumoral, la metástasis y la resistencia al tratamiento. El silenciamiento de CTCFL / BORIS conduce a la senescencia y muerte de CSC. Por lo tanto, una estrategia inmunoterapéutica dirigida a CTCFL / BORIS puede conducir a la destrucción selectiva de CSC y a la posible erradicación de la enfermedad metastásica (Loukinov D; 2018)
CUX1	CUX1 es un gen homeobox que está implicado en la supresión y progresión tumoral (Zubaidah M Ramdzan et al; 2014) Las mutaciones inactivadoras de CUX1 están presentes en muchos tipos de cánceres. El aumento de la expresión de CUX1 es frecuente en muchos tipos de cáncer y se asocia con una supervivencia deficiente. Estas dos funciones opuestas de CUX1 para la patogénesis del cáncer (oncogén y gen supresor de tumores) pueden indicar que varios objetivos transcripcionales y funciones celulares de CUX1, así como el microambiente, afectan la tumorigénesis (Yun Sol Jo et al; 2017)
DDIT3	Proceso biológico: apoptosis, ciclo celular, detención del crecimiento, respuesta al estrés, transcripción, regulación de la transcripción, respuesta de proteínas desplegadas, vía de señalización de Wnt. TF multifuncional en la respuesta al estrés ER. Desempeña un papel esencial en la respuesta a una amplia variedad de tensiones celulares e induce la detención del ciclo celular y la apoptosis en respuesta al estrés ER. Desempeña un doble papel como inhibidor de la función CCAAT / proteína de unión potenciadora (c/EBP) y como activador de otros genes. Desempeña un papel regulador en la respuesta inflamatoria a través de la inducción de caspasa-11 (CASP4 / CASP11) que induce la activación de caspasa-1 (CASP1) y ambas caspasas aumentan la activación de pro-IL1B para IL1B maduro que está involucrado en el respuesta inflamatoria.
DMBX1	Sobre-expresión en tejido (glándula pituitaria, testículo). Este gen codifica un miembro de la subfamilia bicoide de TFs que contienen homeodominio. La proteína codificada actúa como un TF y puede desempeñar un papel en el desarrollo del cerebro y los órganos sensoriales (THPA)

Continúa en la siguiente página

Cuadro 14 – continuación de la página anterior

Factor de transcripción	Descripción
DSTN (ADF-1)	Proteína despolimerizante de actina. Corta los filamentos de actina (F-actina) y se une a los monómeros de actina (G-actina). Actúa de manera independiente del pH (Uniprot)
E2F1	Sobre-expresión en tejido (médula ósea, tejido linfoide). El activador de myc (activación inducida basal + E1A), E2F parece ser un objetivo funcional para la acción de la proteína supresora de tumores Rb, tiene un papel importante en el control de la proliferación celular (Nevins JR; 1992). La sobreexpresión puede causar transformación neoplásica (Singh P et al; 1994). La expresión y función desregulada de E2F1 puede tener consecuencias oncogénicas. Los genes E2F1 y DP1, que codifican parejas heterodiméricas que juntas crean actividad E2F, actúan de manera oncogénica (David G. Johnson et al; 1994). La expresión de E2F1 está regulada positivamente y se correlaciona positivamente con los fenotipos malignos del CCR (Zejun Fang et al; 2019)
E2F4	NRF1, junto con E2F4 y MYC, puede desempeñar un papel crítico en la adquisición de las características del cáncer (Kaumudi Bhawe et al; 2018). Los TFs (ETS1, ESR1, GATA1, GATA2, GATA3, AR, YBX1, FOXP3, E2F4 y PRDM14) se detectaron como moléculas indicadoras de CCR (Md Rezanur Rahman et al; 2019)
E2F5	Juega un papel crucial en el control del ciclo celular y la acción de las proteínas supresoras de tumores y también es un objetivo de las proteínas transformadoras de los virus tumorales de ADN pequeños (THPA).
E2F6	E2F6 podría regular la proliferación, invasión y apoptosis de las células de cáncer gástrico (CG) mediante la inhibición de la expresión de CASC2, lo que sugiere que el eje E2F6/CASC2 es otro regulador de la progresión de dicha enfermedad (Li Y, Jiang L et al; 2019). E2F4 y 6 son nuevos biomarcadores para el pronóstico del cáncer de mama (Sun CC et al; 2019). Papel crucial en el control del ciclo celular (Shi X et al; 2019)
E2F7	Sobre-expresión en tejido (tejido linfoide). Juega un papel esencial en la regulación de la progresión del ciclo celular (Luisa Di Stefano et al; 2003) y en la regulación negativa de un conjunto de miRNA que promueven la proliferación celular (Jone Mitxelena et al; 2016)
EBF1	Grupo enriquecido (células B de memoria, célula B ingenua -aún no expuesta a un antígeno-) Activador transcripcional que reconoce variaciones de la secuencia palindrómica 5'-ATTCCCNNGGGAATT-3 '(Uniprot)
EGR1	Desempeña un papel en la regulación de la supervivencia celular, la proliferación y la muerte celular. Activa la expresión de TP53 y TGFB1, y por lo tanto ayuda a prevenir la formación de tumores; Media las respuestas a la isquemia y la hipoxia; Regula la expresión de proteínas como IL1B y CXCL2 que están involucradas en procesos inflamatorios y el desarrollo de daño tisular después de la isquemia (THPA)

Continúa en la siguiente página

Cuadro 14 – continuación de la página anterior

Factor de transcripción	Descripción
EGR3	Sobre-expresión en tejido (cerebro). EGR 2 y 3 juegan un papel crucial en la regulación del sistema inmune y su participación en la activación de las células B y T, la inducción de anergia y la prevención de la enfermedad autoinmune (Sina Taefehshokre et al; 2017).
EHF	Sobre-expresión en tejido (conducto deferente, glándula salival, vesícula seminal). Activador transcripcional que puede desempeñar un papel en la regulación de la diferenciación y proliferación de células epiteliales. Puede contribuir al desarrollo y la carcinogénesis actuando como un gen supresor de tumores o antoncogen; Marcador pronóstico en CCR (favorable). La sobreexpresión de EHF se correlaciona positivamente con una pobre diferenciación, estadio T avanzado y una supervivencia general más corta de los pacientes con CCR. Los experimentos de función revelaron que la sobreexpresión de EHF promovió la proliferación, migración e invasión de células CCR <i>in vitro</i> e <i>in vivo</i> (Lan Wang et al; 2020)
ELF3	Participa en la mediación de la inflamación vascular. Puede desempeñar un papel importante en la diferenciación de células epiteliales y la tumorigénesis. La regulación aberrante de la vía Wnt / b-catenina juega un papel importante en la carcinogénesis colorrectal, con más del 90 % de los casos de cáncer de colon esporádico con acumulación de b-catenina. Elf3, miembro de la familia E-veintiséis de TFs, impulsa la transactivación de b-catenina y se asocia con una pobre supervivencia de los pacientes con CCR. La supresión de Elf3 en las células CCR atenúa la señalización de b-catenina y disminuye la proliferación celular, la migración y la supervivencia (J-L Wang et al; 2014)
ELK1	Activador. Proteínas relacionadas con la vía RAS (Uniprot). La inhibición de Elk1 indujo la apoptosis de las células de cáncer de tiroides y restringió su proliferación regulando EGR1/PTEN, lo que indica un papel potencial para Elk1 en el tratamiento del cáncer de tiroides (Yakun Kong et al; 2018). La sobreexpresión inducida por Elk1 de TRPM2-AS promovió el desarrollo y la progresión del CG en parte a través del eje de señalización miR-195/HMGA1, y estableció su candidatura como un nuevo biomarcador de cáncer para pacientes con CG. MicroRNA-873 inhibe la metástasis del CCR apuntando a Elk1 y STRN4 (Chuannan Fan et al; 2019)
EN1	Sobre-expresión en tejido (tejido adiposo, cerebro, músculo esquelético, piel). Papel en el control del desarrollo.
EP300	Proteína co-activadora transcripcional celular p300 asociada a E1A, importante en los procesos de proliferación y diferenciación celular. Un co-activador de HIF1A (factor 1 alfa inducible por hipoxia), y por lo tanto juega un papel en la estimulación de genes inducidos por hipoxia (THPA). La proteína de unión a CREB y p300 representan acetiltransferasas de histona y coactivadores transcripcionales que juegan papeles esenciales en la iniciación y progresión del tumor (Rühlmann F et al; 2019)

Continúa en la siguiente página

Cuadro 14 – continuación de la página anterior

Factor de transcripción	Descripción
ESR1	Sobre-expresión en tejido (cuello uterino, uterino, endometrio). Células inmunes mejoradas (células T CD4 ingenuas). Media la regulación génica en respuesta a la hormona estrógeno, con AP-1 / c-Jun, c-Fos, ATF-2, Sp1 y Sp3. Receptor de hormona nuclear. Las hormonas esteroides y sus receptores están involucrados en la regulación de la expresión del gen eucariota y afectan la proliferación y diferenciación celular en los tejidos diana. También puede actuar sinérgicamente con NF-kB para activar la transcripción que involucra elementos de respuesta adyacentes de reclutamiento respectivos; la función involucra CREBBP. (THPA)
ESR2	Sobre-expresión de tejido (ovario, testículo, vagina). El estrógeno está involucrado en la fisiopatología del cáncer de mama, ESR2 puede proteger contra el efecto mitógeno del estrógeno en el tejido mamario (Koyama S et al; 2010). Se une a los estrógenos con una afinidad similar a la de ESR1. Media la regulación génica en respuesta a la hormona estrógeno, con AP-1 / c-Jun, c-Fos, ATF-2, Sp1 y Sp3. Receptor de hormona nuclear. Las hormonas esteroides y sus receptores están involucrados en la regulación de la expresión génica eucariota y afectan la proliferación y diferenciación celular en los tejidos diana (THPA) (Paulo Roberto Stevanato Filho; 2018)
ETS1	Sobre-expresión de tejido (tejido linfático). Actúa sinérgicamente con AP-1, SRF y PAX (Verger et al., 2002). Se cree que se une a TP53, lo que le confiere un papel supresor de tumores (Fry et al., 2018). La desregulación del TF facilita la proliferación celular en los cánceres, y varios miembros participan en la invasión y metástasis activando la transcripción génica. Participa en la oncogénesis y la supresión tumoral según las situaciones biológicas utilizadas (Elizabeth A. Fry y Kazushi Inoue; 2018). Participa en el desarrollo de células madre, senescencia y muerte celular y tumorigénesis (Uniprot)
ETV1	Sobre-expresión de tejido (cerebro, glándula salival). ETV1 es crítico en la expansión del estroma desmoplásico y la progresión metastásica del cáncer de páncreas en ratones, mediado funcionalmente en parte a través de Sparc y Has2 (Steffen Heeg et al; 2016). Poca expresión en CCR (Gepia)
ETV4	Sobre-expresión de tejido (vesícula biliar). Activador transcripcional que se une al potenciador del gen adenovirus E1A. Actúa como mediador de metástasis, y proporciona una justificación biológica para el pronóstico adverso asociado con la expresión elevada de ETV4 en el CCR humano (Aruz Mesci et al; 2014). El eje CIC-ETV4 no es solo un módulo clave que controla la progresión del CCR, sino también un nuevo objetivo terapéutico y/o diagnóstico para el CCR (Jeon-Soo Lee et al; 2020).
ETV5	Sobre-expresión de tejido (glándula paratiroides). Los TFs de ETS juegan papeles importantes en la invasión, diferenciación y angiogénesis de células tumorales. El factor de transcripción ETV5 podría estimular la malignidad de CCR y promover la angiogénesis de CCR al apuntar directamente a PDGF-BB

Continúa en la siguiente página

Cuadro 14 – continuación de la página anterior

Factor de transcripción	Descripción
ETV6	Este gen codifica un TF de la familia ETS. El producto de este gen contiene dos dominios funcionales: un dominio N-terminal puntiagudo que participa en las interacciones proteína-proteína consigo mismo y otras proteínas, y un dominio de unión al ADN C-terminal. Los estudios de genes inactivados en ratones sugieren que se requiere para la hematopoyesis y el mantenimiento de la red vascular en desarrollo. Se sabe que este gen está involucrado en una gran cantidad de reordenamientos cromosómicos asociados con leucemia y fibrosarcoma congénito (THPA)
EZH2	Sobre-expresión de tejido (médula ósea, tejido linfoide). EZH2 controla reguladores del desarrollo en células madre embrionarias y es esencial para la determinación y transición del destino celular. EZH2 se informó sobre-expresado en una variedad de tumores sólidos. EZH2 reprime la expresión de una gran cantidad de genes supresores de tumores en las células tumorales, lo que promueve el ciclo celular, la proliferación celular y la invasión celular y la progresión del cáncer. Recientemente, está surgiendo evidencia que indica roles importantes de EZH2 en las células inmunes. EZH2 juega un papel en la regulación de varios tipos de células inmunes, el microambiente tumoral, las respuestas inmunes y las inmunoterapias contra el cáncer (Xiaohai Wang et al; 2019).
EWSR1	Este gen codifica una proteína multifuncional que participa en diversos procesos celulares, incluida la expresión génica, la señalización celular y el procesamiento y transporte de ARN (THPA). EWSR1-CREB1 y EWSR1-ATF1 son fusiones de genes descritas consistentemente en 5 neoplasias histopatológicas y conductualmente diversas (entre ellas encontramos tumores del tracto gastrointestinal) (Khin Thway et al; 2012)
FLI1	Sobre-expresión en tejido (tejido linfoide). FLI1 es un nuevo controlador oncogénico que promueve la metástasis tumoral a través de la vía miR584-ROCK1. Es importante destacar que el FECR1 exosómico en suero puede servir como un biomarcador prometedor para rastrear la progresión de la enfermedad de SCLC (cáncer de pulmón de células pequeñas). Casi sin expresión en CCR (Gepia)
FOS	Gen relacionado con el cáncer. Sinónimo: AP1; Fos / Jun; PEA1. Protooncogen. Papel dominante de AP-1 en la regulación positiva transcripcional de MMP-1 en células CCR que está modulada por las funciones conjuntas de STAT3 y STAT1 (Muller A1,2, Gasch J et al; 2019). Involucrado en la vía MAPK, tiene un papel clave en la inflamación, proliferación, diferenciación y apoptosis regulada por estímulos externos, incluida la infección viral (Gazon et al; 2018). c-Myb promueve el crecimiento y la metástasis del CCR a través de la transición epitelial-mesenquimal inducida por c-fos (Xiao Qu et al; 2019)
FOSB, FOSL1, FOSL2	La familia de genes Fos consta de 4 miembros: FOS, FOSB, FOSL1 y FOSL2. Estos genes codifican proteínas de cremallera de leucina que pueden dimerizarse con proteínas de la familia JUN, formando así el complejo del TF AP-1. Como tal, las proteínas FOS han sido implicadas como reguladores de la proliferación, diferenciación y transformación celular (THPA).

Continúa en la siguiente página

Cuadro 14 – continuación de la página anterior

Factor de transcripción	Descripción
FOXA1	Sobre-expresión en tejido (hígado, próstata). FOXA1 ha sido atribuido previamente a roles supresores de tumores y oncogénicos. Sobreexpresión en adenocarcinoma de próstata y carcinoma invasivo de mama, no hay expresión en CCR (Gepia)
FOXD3	CEP55, FOXD3, FOXF2, GNAO1, GRIA4 y KCNA5 expresados diferencialmente en CCR en comparación con la mucosa normal y cuya expresión parece depender de la metilación. Todos estos seis genes eran comunes en todos los grupos de metilación y estados de mutación de CCR y, como tal, se cree que es un evento temprano en la carcinogénesis humana de CCR (Nina Hauptman et al; 2019). La restauración de la expresión de FOXD3 inhibió la proliferación y migración de las células tumorales. FOXD3, que con frecuencia se metila en el CCR, actuó como un supresor tumoral que induce la apoptosis de las células tumorales bajo estrés ER a través de p53 (Ming Xu; 2019). De acuerdo con THPA y Gepia, existe poca o ninguna expresión de FOXD3 en el CCR, sin embargo, los documentos y las revisiones muestran muchos casos en los que FOXD3 podría usarse como biomarcador en CCR.
FOXM1	Sobre-expresión de tejido (tejido linfoide, testículo). Sobreexpresión en CCR (Gepia). Varias proteínas (NPM, MELK, Pin1, STAT3, NF-kB / p65, CENPF) estabilizan FOXM1 y aumentan su actividad oncogénica (Andrei L. Gartel et al; 2017). Elemental para la regulación de GLI1.
FOXO1	Sobre-expresión en tejido (músculo esquelético). TF que es el objetivo principal de la señalización de la insulina y regula la homeostasis metabólica en respuesta al estrés oxidativo. Muy poca expresión en CCR (Gepia).
FOXO3	Desencadena la apoptosis en ausencia de factores de supervivencia, incluida la muerte celular neuronal por estrés oxidativo (A Brunet et al; 1999). Participa en la regulación postranscripcional de MYC (THPA)
FOXP3	Sobre-expresión en tejido (sangre, tejido linfoide). Crítico para el control de las respuestas inmunes y posiblemente el número de células T, actuando como un regulador para la activación de las células T (Brunkow ME et al; 2001)
GABPA	TF capaz de interactuar con repeticiones ricas en purinas (repeticiones GA). Necesario para la expresión del gen Adenovirus E4.
GCFC2	Factor que reprime la transcripción. Se une a las secuencias ricas en GC (5'-GCCGGGC-3') presentes en el receptor del factor de crecimiento epidérmico, la beta-actina y los promotores de proteasa dependientes de calcio (Uniprot).
GCM1	Sobre-expresión de tejido (placenta). El TF involucrado en el control de la expresión del factor de crecimiento placentario y otros genes específicos de placenta (K Yamada et al; 1999) (Miao Chang et al; 2008) tiene un papel central en la mediación de la diferenciación de las células del trofoblasto a lo largo de ambas las vías vellosas y extravillanas en el desarrollo placentario (D Baczyk et al; 2009)
GLI2	Sobre-expresión en tejido (ovario). Funciona como regulador de la transcripción en la ruta del erizo (Hh) (Markku Varjosalo ET AL; 2008) (Bingchen Han et al; 2015). Funciona como activador transcripcional (A Tanimura et al; 1998)

Continúa en la siguiente página

Cuadro 14 – continuación de la página anterior

Factor de transcripción	Descripción
GLIS2	Puede actuar como represor transcripcional o como activador transcripcional, dependiendo del contexto celular. Desempeña un papel en el desarrollo renal y la neurogénesis (Feng Zhang et al; 2002)
GRB2	Proceso biológico: interacción virus-huésped (Uniprot). Los 14 genes (incluidos BCAR1, CRK, FYN, GRB2, LCP2, PIK3R1, PLCG1, PTK2, PTPN11, PTPN6, SHC1, SOS1, SRC y SYK) en este módulo son genes inflamatorios, enfatizando el papel crítico de la inflamación en el CCR (Hao Jiang et al; 2018).
HBP1	Represor transcripcional que se une a la región promotora de genes diana. Desempeña un papel en la regulación del ciclo celular y de la vía Wnt. Interrumpe la interacción entre el ADN y TCF4 (Uniprot)
HES1	Puede actuar como un regulador negativo de la miogénesis al inhibir las funciones de MYOD1 y ASH1 (Uniprot)
HIC1	Crecimiento regulador y gen represor tumoral. Participa en la regulación negativa de SIRT1 y, por lo tanto, participa en la regulación de las respuestas de daño de ADN apoptótico dependiente de p53/TP53. La IL-6 regula a la baja el supresor tumoral HIC1 y promueve el desarrollo del cáncer de mama en el microambiente tumoral mediante señalización paracrina o autocrina (Sun X et al; 2019). La proteína desacetilasa SIRT1 es crucial para numerosos procesos fisiológicos, como el envejecimiento, el metabolismo y la autoinmunidad, y está reprimido por varios TFs, incluido HIC1 (Song JY et al; 2019).
HMGB1	Proceso biológico: inmunidad adaptativa, autofagia, quimiotaxis, daño en el ADN, recombinación de ADN, reparación de ADN, interacción virus-huésped, inmunidad, respuesta inflamatoria, inmunidad innata. Promueve la respuesta inflamatoria del huésped a las señales estériles e infecciosas y participa en la coordinación e integración de las respuestas inmunes innatas y adaptativas. En el citoplasma funciona como sensor y / o chaperona para los ácidos nucleicos inmunogénicos que implican la activación de las respuestas inmunes mediadas por TLR9, y media la autofagia. Actúa como molécula de patrón molecular asociado al peligro (DAMP) que amplifica las respuestas inmunes durante la lesión tisular (Daphne C Avgousti et al; 2016)
HNF1A	Sobre-expresión de tejido (intestino, riñón, hígado). La proteína codificada por este gen es un TF requerido para la expresión de varios genes específicos del hígado. HNF1A-AS1 se reguló significativamente tanto en los tejidos tumorales CCR como en las líneas celulares CCR en comparación con los tejidos adyacentes no tumorales y la línea celular epitelial del colon humano normal (HcoEpiC). La disminución de la expresión de lncRNA HNF1A-AS1 suprimió la actividad de la vía de señalización de Wnt / b-catenina al regular negativamente la expresión de b-catenina, cyclinD1 y c-myc (Xin Zhang et al; 2017).
Continúa en la siguiente página	

Cuadro 14 – continuación de la página anterior

Factor de transcripción	Descripción
HNF1B	Sobre-expresión en tejido (vesícula biliar, riñón, páncreas). La proteína se une al ADN como un homodímero o como un heterodímero con la proteína relacionada con el HNF1A. Se ha demostrado que el gen funciona en el desarrollo de nefronas y regula el desarrollo del páncreas embrionario. Las mutaciones en este gen provocan quistes renales y síndrome de diabetes y diabetes mellitus no dependiente de insulina, y la expresión de este gen está alterada en algunos tipos de cáncer (Uniprot, THPA)
HNF4A	Sobre-expresión en tejido (intestino, riñón, hígado). HNF4A controla la homeostasis del epitelio intestinal y la absorción intestinal de los lípidos de la dieta (Floriane Baraille et al; 2015). Actúa como un activador transcripcional para la expresión génica de eritropoyetina específica de tejido e hipoxia (Galson DL et al; 1995). HNF4A regula el equilibrio intestinal entre la proliferación y la diferenciación (Anne-Laure Cattin et al; 2009)
HOXA3	La expresión de HoxA3 en estas células afecta la formación adecuada de los órganos de timo, tiroides y paratiroides. La eliminación de HoxA3 conduce al fracaso en la formación del timo y la glándula paratiroides. HoxA3 mutante también causa una reducción en el tamaño de la tiroides. Si bien las células foliculares y parafoliculares aún se diferencian, sus números se reducen y no se distribuyen uniformemente en toda la glándula (Manley NR et al, 1995; Chojnowski JL et al, 2014)
HOXA13	Sobre-expresión en tejido (cuello uterino, uterino, próstata). Regulación positiva de HOXA13 y la progresión del CG. Además, HOXA13 contribuye a la invasión y EMT de las células de CG a través de la vía de señalización de TGF- β (Y-X He et al; 2017). La sobreexpresión de HOXA13 promueve la proliferación y la metástasis en parte a través de la activación de Erk1 / 2 en CG (Zhiwei Qin et al; 2019).
HOXB13	Sobre-expresión en tejido (intestino, próstata). TF específico de secuencia que forma parte de un sistema regulador del desarrollo que proporciona a las células identidades posicionales específicas en el eje anterior-posterior. Se une preferentemente al ADN metilado (Yimeng Yin et al; 2017)
HOXD10	Sobre-expresión de tejido (mama, endometrio). TF específico de secuencia que forma parte de un sistema regulador del desarrollo que proporciona a las células identidades posicionales específicas en el eje anterior-posterior (Uniprot)
HOXD9	Sobre-expresión de tejido (endometrio). HOXD9 está involucrado en el crecimiento del cáncer y la metástasis. La expresión ectópica de HOXD9 promovió la metástasis de invasión en las células del CCR mediante la inducción de EMT <i>in vitro</i> e <i>in vivo</i> (Mengwei Liu et al; 2020)
IKZF1	Sobre-expresión en tejido (sangre, médula ósea, tejido linfoide). Regulador de la transcripción de la diferenciación de células hematopoyéticas, funciona como un regulador de la diferenciación de linfocitos (Marilyne Dijon et al; 2008). Desempeña un papel en el desarrollo de linfocitos, células B y T. La sobreexpresión de algunas isoformas negativas dominantes se ha asociado con neoplasias malignas de células B, como la leucemia linfoblástica aguda (Uniprot, THPA)

Continúa en la siguiente página

Cuadro 14 – continuación de la página anterior

Factor de transcripción	Descripción
INSM1	Sobre-expresión de tejido (glándula suprarrenal, cerebro, glándula pituitaria). Participa en el ciclo celular, diferenciación, neurogénesis, transcripción, regulación de la transcripción. La expresión de INSM1 aumentó significativamente en tejido neoplásico versus no neoplásico. Las neoplasias con metástasis conocidas mostraron una expresión significativamente mayor que las que aún no habían hecho metástasis (Rosenbaum JN et al; 2015). Marcador sensible para la diferenciación neuroendocrina de tumores pulmonares humanos (Gepia)
IRF1	Sobre-expresión de tejido (sangre). La proteína codificada por este gen es un regulador transcripcional y un supresor tumoral, que sirve como un activador de genes involucrados en las respuestas inmunes innatas y adquiridas. La proteína codificada activa la transcripción de genes involucrados en la respuesta del cuerpo a virus y bacterias, desempeñando un papel en la proliferación celular, la apoptosis, la respuesta inmune y la respuesta al daño del ADN. Esta proteína reprime la transcripción de varios otros genes. Como supresor tumoral, suprime el crecimiento de células tumorales y estimula una respuesta inmune contra las células tumorales. Los defectos en este gen se han asociado con CG, leucemia mielógena y cáncer de pulmón (THPA).
ITGAM	Sobre-expresión de tejido (médula ósea). Proceso biológico: adhesión celular, inmunidad, inmunidad innata. Integrin ITGAM / ITGB2 está implicado en diversas interacciones adhesivas de monocitos, macrófagos y granulocitos, así como en la mediación de la absorción de partículas y patógenos recubiertos de complemento (Josephine Losse et al; 2010). Regula la migración de neutrófilos (Ming Bai et al; 2017)
JDP2	Proteína de dimerización de Jun 2. Componente del TF AP-1 que reprime la transactivación mediada por la familia de proteínas Jun. Participa en una variedad de respuestas transcripcionales asociadas con AP-1, tales como apoptosis inducida por UV, diferenciación celular, tumorigénesis y antitumogeneris (Uniprot)
c-JUN	Protooncogen. Participa en la activación transcripcional mediada por KRAS activada de USP28 en células de CCR. Se une al promotor USP28 en células de CCR (Ryan W Serra ET AL; 2014). Codifica una proteína que es muy similar a la proteína viral y que interactúa directamente con secuencias específicas de ADN objetivo para regular la expresión génica.
JUNB	JunB es miembro de la familia AP-1 de TFs diméricos. Ejerce una doble acción en el ciclo celular. Es mejor conocido como un inhibidor de la proliferación celular, un inductor de senescencia y un supresor tumoral (Marc Piechaczyk et al .; 2008)
JUND	Sobre-expresado en tejido (médula ósea). La proteína codificada por este gen es un miembro de la familia JUN y un componente funcional del complejo del TF AP1. Esta proteína se ha propuesto para proteger las células de la senescencia y la apoptosis dependientes de p53 (RefSeq, noviembre de 2013)

Continúa en la siguiente página

Cuadro 14 – continuación de la página anterior

Factor de transcripción	Descripción
KLF1	Sobre-expresión en tejido (médula ósea, tejido linfoide, placenta). La activación de KLF1 mejora la diferenciación y maduración de los glóbulos rojos de las células madre pluripotentes humanas (Cheng - Tao Yang ET AL; 2017). Klf1 y Egr2 son moduladores de la supresión inmune mediada por PD-L1 en las células T CD4+ y podrían proporcionar nuevas ideas sobre objetivos terapéuticos para enfermedades autoinmunes y tumores malignos (Shuzo Teruya et al; 2018). KLF1 es un importante TF eritroide que une principalmente potenciadores y supereforzadores para activar la expresión génica (Melissa D. Ilsley et al; 2017)
KLF4	Sobre-expresión en tejido (piel). Correlación inversa entre la expresión de miR-543 y KLF4 en tejidos de CCR. La regulación negativa de KLF4 favorece el efecto oncogénico inducido por miR-543 sobre la proliferación celular, la apoptosis, la migración y la invasión (Zhai F et al; 2017). MiR-543 facilita la proliferación y metástasis del CCR al enfocarse en KLF4 (Babaei K et al; 2019). KLF4 se encontró regulado negativamente en CCR (Martha L. Slattery1 et al; 2017)
KLF5	Sobre-expresión en tejido (esófago, intestino). La expresión y la actividad de KLF5 están reguladas por múltiples vías de señalización, incluidas Ras / MAPK, PKC y TGFb, y varias modificaciones postraduccionales. KLF5 media las funciones de señalización en la proliferación celular, ciclo celular, apoptosis, migración, diferenciación y potencia mediante la regulación de la expresión génica en respuesta a estímulos ambientales. El ácido lisofosfatídico estimula la proliferación celular al inducir la expresión de KLF5 en líneas celulares de cáncer de colon. KLF5 tiene una función anti-apoptosis, puede modular la apoptosis de una manera independiente de p53. KLF5 promueve la migración de células epiteliales (Jin-Tang Dong et al, 2009; Zhang H et al, 2007)
KLF6	La disminución de la expresión de KLF6-SV2 puede estar asociada con la aparición y el desarrollo de CCR. KLF6-SV2 desempeña un papel como supresor tumoral al bloquear eficazmente la proliferación celular, detener el ciclo celular e inducir la apoptosis en el CCR, lo que puede estar relacionado con una mayor expresión de p21 y Bax (Biao Zhang et al; 2018)
KLF9	TF que se une a los elementos promotores de la caja GC. Activa selectivamente la síntesis de ARNm a partir de genes que contienen repeticiones en tándem de cajas GC pero reprime los genes con una sola caja GC. Actúa como un TF circadiano epidérmico que regula la proliferación de queratinocitos (Uniprot). KLF9 suprime significativamente la invasión de células GC y la metástasis al inhibir la transcripción de MMP28 (Yang Li et al; 2019); KLF9 exhibió baja expresión en cáncer de páncreas, y la regulación positiva de KLF9 puede inhibir la progresión del cáncer de páncreas (Zhiwei Zhong et al; 2018)
LEF1	Sobre-expresión de tejido (tejido linfoide). Participa en la vía de señalización Wnt. Activa la transcripción de genes diana en presencia de CTNNB1 y EP300. La isoforma 1 activa transcripcionalmente la expresión de MYC y CCND1 y aumenta la proliferación de células tumorales pancreáticas (Uniprot)

Continúa en la siguiente página

Cuadro 14 – continuación de la página anterior

Factor de transcripción	Descripción
LHX1	Sobre-expresión de tejido(cerebro, riñón, vesícula seminal). Sin expresión en CCR según Gepia
MAF	Activa la expresión de IL4 en células T helper 2 (Th2). Aumenta la susceptibilidad de las células T a la apoptosis al interactuar con MYB y disminuir la expresión de BCL2. Junto con PAX6, transactiva fuertemente el promotor del gen glucagón a través del elemento G1. Cuando se sobreexpresa, reprime la transcripción mediada por el elemento de respuesta antioxidante (ARE). Participa como oncogén o como supresor tumoral, según el contexto celular (Uniprot).
MAFB	Los niveles de MAFB aumentaron en muestras clínicas de CCR, y una mayor expresión se correlacionó con una etapa TNM más avanzada (Lin-Sen Yang et al; 2016)
MAZ	Puede funcionar como un TF con roles duales en el inicio y la terminación de la transcripción. Se une a dos sitios, ME1a1 y ME1a2, dentro del promotor MYC que tiene mayor afinidad por el primero. También se une a múltiples sitios ricos en G/C dentro del promotor de la familia Sp1 de factores de transcripción. Regula la expresión inducida por inflamación de las proteínas amiloides A séricas (Uniprot)
MECOM	Funciona como un regulador transcripcional que se une a las secuencias de ADN en la región promotora de los genes diana y regula positiva o negativamente su expresión. Oncogén que juega un papel en el desarrollo, la proliferación y diferenciación celular. También puede desempeñar un papel en la apoptosis a través de la regulación de la señalización JNK y TGF- β . Implicado en la hematopoyesis (Uniprot)
MEF2A	Interviene en las funciones celulares de desarrollo del músculo esquelético y cardíaco, en la diferenciación y supervivencia neuronal. Desempeña diversos roles en el control del crecimiento celular, la supervivencia y la apoptosis a través de la señalización de p38 MAPK en la transcripción específica del músculo y / o relacionada con el factor de crecimiento (Uniprot).
MEIS2	Participa en la regulación transcripcional. Se une a las proteínas HOX o PBX para formar dímeros y se cree que tiene un papel en la estabilización del complejo homeoproteína-ADN. Probablemente en complejo con PBX1, está involucrado en la regulación transcripcional por KLF4 (THPA)
MNX1	Sobre-expresión de tejido (páncreas). La expresión ectópica de MNX1 activa la señalización de Wnt / β -catenina y aumenta la expresión de c-Myc y CCND1, los genes posteriores de la señalización de Wnt / β -catenina. Por lo tanto, MNX1 desempeña un papel indispensable en la promoción de la progresión del CCR humano y puede representar un nuevo objetivo terapéutico para el CCR (Xiangrong Yang et al; 2019). Sobreexpresión en CCR según Gepia.

Continúa en la siguiente página

Cuadro 14 – continuación de la página anterior

Factor de transcripción	Descripción
MSX1	Sobre-expresión de tejido (cuello uterino). Actúa como represor transcripcional. Involucrado en múltiples interacciones epiteliales-mesenchimales durante la embriogénesis de vertebrados. Las variantes de MSX1 se han relacionado con la agenesia dental, la hendidura orofacial y la displasia ungueal. Participa en la formación de extremidades, el desarrollo del sistema nervioso y la inhibición del crecimiento tumoral (Jia Liang et al; 2016)
MTF1	Activa el promotor de metalotioneína I, se une al elemento sensible al metal (MRE) (Uniprot, THPA)
MTUS1	La isoforma 1 inhibe la proliferación celular del cáncer de mama, retrasa la progresión de la mitosis al prolongar la metafase y reduce el crecimiento tumoral (Uniprot)
MYB	Sobre-expresión de tejido (tejido linfóide). Los componentes estructurales de v-myb son importantes para la transactivación de la expresión génica (Bortner DM1, Ostrowski MC.; 1991). La proteína de control transcripcional, caracterizada por un dominio de unión al ADN N-terminal altamente conservado, se considera un oncogén (Klempnauer KH1, Sippel AE; 1987). c-Myb se sobreexpresa significativamente en los tejidos de CCR en comparación con los tejidos normales adyacentes. La alta expresión de c-Myb se correlaciona positivamente con metástasis de ganglios linfáticos y mal pronóstico. c-Myb inhibe la proliferación, resistencia a la apoptosis, invasión, metástasis, formación de colonias y tumorigénesis <i>in vivo</i> de células CCR. Además, la investigación del mecanismo indica que c-Myb puede promover la progresión del CCR mediante la regulación de c-fos. La sobreexpresión de c-fos puede rescatar el efecto inhibitorio de c-Myb sobre las características malignas de las células de CCR: c-Myb inhibe el fenotipo molecular de EMT en células de CCR, mientras que la sobreexpresión de c-fos puede rescatar este efecto inhibitorio. Este estudio sugiere que c-Myb promueve la progresión maligna de CCR a través de EMT inducida por c-fos y tiene el potencial de ser un biomarcador pronóstico prometedor y un objetivo terapéutico. (Xiao Qu et al; 2019)
MYC	Protooncogen que juega un papel en la progresión del ciclo celular, la apoptosis y la transformación celular. La proteína codificada forma un heterodímero con el TF relacionado MAX. Este complejo se une a la secuencia consenso de ADN de la caja E y regula la transcripción de genes diana específicos. La amplificación de este gen se observa con frecuencia en numerosos cánceres humanos (THPA). La capacidad de Myc para activar la telomerasa puede contribuir a su capacidad de promover la formación de tumores (Wang J et al; 1998). Regulada negativamente por AP-2 (Gaubatz S et al; 1995). Puede estar involucrado en la regulación postranscripcional del metabolismo de rRNA, lo que podría explicar, en parte, la estimulación generalizada de las células durante la tumorigénesis

Continúa en la siguiente página

Cuadro 14 – continuación de la página anterior

Factor de transcripción	Descripción
MYOD1	Sobre-expresión de tejido (músculo esquelético). Activador de la transcripción muscular específica (D Sassoon et al; 1989); se observó hipermetilación de este gen en comparación con la mucosa normal en el 88 % de los adenomas y en el 99 % de los carcinomas. La metilación extensa fue significativamente más frecuente en tumores con inestabilidad de microsatélites (B J Iacopetta et al, 1997; B Shannon et al, 1999). Según Gepia, no hay expresión en tejido colorrectal normal ni necrótico.
MYOG	Sobre-expresión de tejido (músculo esquelético). Activador de la transcripción muscular específico que puede inducir miogénesis en una variedad de tipos de células (D Sassoon et al; 1989); Esencial para el desarrollo de la diferenciación funcional del músculo de fibra esquelética embrionaria y para el músculo esquelético funcional [RefSeq, Jul 2008].
MZF1	Se une al ADN del promotor objetivo y funciona como regulador de la transcripción. Regula la transcripción del promotor PADI1 y CDH2. Puede ser un regulador de los eventos transcripcionales durante el desarrollo hemopoyético (THPA).
NEUROD1	Contribuye a la regulación de varias vías de diferenciación celular, como las que promueven la formación de células ganglionares retinianas tempranas, neuronas sensoriales del oído interno, células granulares que forman el cerebelo o la capa de células de giro del hipocampo, células de los islotes endocrinos del páncreas y células enteroendocrinas del intestino delgado. También se requiere para la morfogénesis dendrita y el mantenimiento en la corteza cerebelosa. Se asocia con la cromatina a los elementos reguladores potenciadores en genes que codifican reguladores transcripcionales clave de la neurogénesis (Uniprot)
NF1	Funciona como un regulador negativo de la vía de transducción de señales ras. Se considera un supresor tumoral porque la reducción de su producción (en células haploinsuficientes) implica la activación de la proteína ras, que a su vez regula una cascada de vías de señalización posteriores, incluidas las proteínas quinasas activadas por mitógeno (MAPK), P13K, PKB y proteína quinasa diana de rapamicina en células de mamífero (mTOR) 3. La activación de estas vías causa una variedad de efectos celulares que generalmente estimulan la supervivencia y proliferación celular (E Denayer et al; 2008)
NFATC1	Desempeña un papel en la expresión inducible de genes de citoquinas en células T, especialmente en la inducción de la transcripción del gen IL-2 o IL-4. También controla la expresión génica en las células cardíacas embrionarias. Podría regular no solo la activación y la proliferación, sino también la diferenciación y la muerte programada de los linfocitos T, así como de las células linfoides y no linfoides (S Chuvpilo et al; 1999). Se requiere para la osteoclastogénesis y regula muchos genes importantes para la diferenciación y función de los osteoclastos.

Continúa en la siguiente página

Cuadro 14 – continuación de la página anterior

Factor de transcripción	Descripción
NFATC2	Sobre-expresión de tejido (tejido linfoide). Desempeña un papel en la expresión inducible de genes de citoquinas en células T, especialmente en IL-2, IL-3, IL-4, TNF-alfa o GM-CSF. Promueve la migración invasiva a través de la activación de la expresión GPC6 y la vía de señalización WNT5A (Uniprot). Este complejo juega un papel central en la inducción de la transcripción génica durante la respuesta inmune.
NFATC3	Actúa como regulador de la activación transcripcional. Desempeña un papel en la expresión inducible de genes de citoquinas en células T, especialmente en la inducción de la IL-2 (Guo-Dong Li et al; 2008). Junto con NFATC4, involucrado en el desarrollo del corazón embrionario
NFATC4	TF regulado que participa en varios procesos, incluido el desarrollo y la función del sistema inmunitario, cardiovascular, musculoesquelético y nervioso. Participa en la activación de células T, estimulando la transcripción de genes de citocinas, incluida la de IL2 e IL4. Junto con NFATC3, participa en el desarrollo del corazón embrionario (THPA, Uniprot).
NFE2L1	Sobre-expresión de tejido (músculo esquelético). Sensor de membrana del retículo endoplásmico que se transloca en el núcleo en respuesta a diversas tensiones para actuar como TF (Senthil K Radhakrishnan et al; 2014). Capaz de detectar diversas tensiones celulares, como el exceso de colesterol, el estrés oxidativo o la inhibición del proteasoma (Janos Steffen et al; 2010)
NFE2L2	TF que desempeña un papel clave en la respuesta al estrés oxidativo: se une a los elementos de respuesta antioxidante (ARE) presentes en la región promotora de muchos genes citoprotectores, como las enzimas desintoxicantes de fase 2, y promueve su expresión, neutralizando así los electrófilos reactivos (Peter Huppke et al; 2017) (Viraj R Sanghvi et al; 2019)
NFkB1	Presente en casi todos los tipos de células y es el punto final de una serie de eventos de transducción de señales que son iniciados por una amplia gama de estímulos relacionados con muchos procesos biológicos como inflamación, inmunidad, diferenciación, crecimiento celular, tumorigénesis y apoptosis. NFkB es un factor importante en la patogénesis de CCR (Moorchung N et al; 2014), esencial para la migración del CCR inducida por activina a través de la regulación positiva de la vía PI3K-MDM2 (Jana A et al; 2017). Comúnmente observado en cánceres humanos y provoca la liberación de citocinas, incluida la IL-6 (Yannick Verhoevena et al; 2019). NF-kB es un TF que pertenece a la familia Rel. Se ha demostrado que la activación de NF-kB desempeña un papel en la proliferación celular, la apoptosis, la producción de citocinas y la oncogénesis (Ram Rattan Negi et al., 2018).

Continúa en la siguiente página

Cuadro 14 – continuación de la página anterior

Factor de transcripción	Descripción
NFYA	Conocido como CBF; CBF-A + CBF-B; Factor de unión a CCAAT; CP1; NF-Y; TF heterotrimérico y altamente que se une a factores CCAAT en las regiones promotoras (THPA). Los factores CCAAT son proteínas de unión al potenciador (c/EBP) y están implicados en diferentes respuestas celulares, como en el control de la proliferación celular, el crecimiento y la diferenciación, en el metabolismo y en la inmunidad (Martyn GE et al; 2017). Su susceptibilidad a las mutaciones especifica que la caja CCAAT juega un papel importante en la determinación de la eficiencia del promotor. No parece jugar un papel directo en la especificidad del promotor, pero su inclusión aumenta la fuerza del promotor.
NKX2-1	Sobre-expresión de tejido (pulmón, glándula tiroides). Activa el promotor de genes específicos de tiroides, como tiroglobulina, tiroperoxidasa y receptor de tirotropina. Crucial en el mantenimiento del fenotipo de diferenciación tiroidea, puede estar asociado con el cáncer de tiroides (THPA)
NKX2-5	Sobre-expresión de tejido (músculo del corazón). Funciones en la formación y desarrollo del corazón, las mutaciones causan defecto del tabique auricular con defecto de conducción auriculoventricular (THPA). Expresión diferencial en el glioblastoma multiforme (Gepia)
NKX6-2	Sobre-expresión de tejido (conducto deferente, músculo cardíaco, tejido linfoide, glándula salival). Actúa como un activador transcripcional (Kirsten Heathcote et al; 2005). En combinación con NKX2-5, puede desempeñar un papel en el desarrollo embrionario faríngeo y cardíaco (Uniprot)
NR1H2	Juega funcionalmente con RORA para la regulación de genes implicados en el metabolismo hepático. Marcador pronóstico en CCR (desfavorable)
NR1H3	Sobre-expresión de tejido (tejido adiposo). Receptor nuclear que exhibe una actividad de activación transcripcional dependiente de ligando (Ayako Sakurabashi et al; 2015). Desempeña un papel importante en la regulación de la homeostasis del colesterol, regulando la absorción de colesterol a través de la ubiquitinación dependiente de MYLIP de LDLR, VLDLR y LRP8 (Tomas Jakobsson et al; 2009). Juega funcionalmente con RORA para la regulación de genes implicados en el metabolismo hepático (Uniprot)
NR1H4	Sobre-expresión de tejido(hígado). Regula la homeostasis de los lípidos y la glucosa y está implicado en la respuesta inmune innata. Regula hacia abajo la expresión de citocinas inflamatorias en varios tipos de células inmunes, incluidos los macrófagos y las células mononucleares (Raffaella M Gadaleta et al; 2011)

Continúa en la siguiente página

Cuadro 14 – continuación de la página anterior

Factor de transcripción	Descripción
NR1I2	Sobre-expresión de tejido (sangre, intestino, hígado). Sobreexpresado en CCR (Gepia). Receptor nuclear que se une y se activa por una variedad de compuestos endógenos y xenobióticos. TF que activa la transcripción de múltiples genes involucrados en el metabolismo y la secreción de xenobióticos, drogas y compuestos endógenos potencialmente dañinos. Activado por el antibiótico rifampicina y varios metabolitos vegetales, como la hiperforina, guggulípidos, colupulona e isoflavonas (Uniprot). No se encontró ninguna explicación sobre la relación entre la sobreexpresión del TF y la enfermedad de CCR.
NR2C2	Desempeñar un papel esencial en el desarrollo, la homeostasis celular y la enfermedad, incluido el cáncer, donde la sobreexpresión o la subexpresión de algunos receptores tiene importancia pronóstica para la supervivencia del paciente. Su expresión es un factor pronóstico positivo (cáncer de ovario) y negativo (próstata y mama) para el cáncer (Stephen Safe et al; 2014)
NR2F2	TF activado por ligando. Activado por altas concentraciones de ácido 9-cis-retinoico y ácido todo-trans-retinoico, pero no por dexametasona, cortisol o progesterona. Factor sobre-expresado en célula inmunitaria (memoria de células T CD8) (Uniprot)
NR3C1	Proceso biológico: apoptosis, ciclo celular, división celular, partición cromosómica, mitosis, transcripción, regulación de la transcripción (Uniprot). Este gen codifica el receptor de glucocorticoides, que puede funcionar tanto como un TF que se une a los elementos de respuesta a glucocorticoides en los promotores de los genes sensibles a glucocorticoides para activar su transcripción, así como un regulador de otros TFs. El receptor se encuentra en el citoplasma. Está implicado en respuestas inflamatorias, proliferación celular y diferenciación en tejidos diana. Las mutaciones en este gen están asociadas con la resistencia a los glucocorticoides generalizada (THPA)
NR4A2	Sobre-expresión de tejido (vesícula seminal). La alta expresión de NR4A2 en las células de CCR confiere resistencia a la quimioterapia, atenúa la apoptosis inducida por la quimioterapia y predice un pronóstico desfavorable de los pacientes con cáncer de colon, especialmente para aquellos que recibieron quimioterapia postoperatoria. NR4A2 puede ser un factor pronóstico y predictivo para el CCR (Yifang Han et al; 2013). La alta expresión de NR4A2 en las células GC confiere quimiorresistencia, atenúa la apoptosis inducida por 5-FU y predice una supervivencia desfavorable, especialmente para aquellos que recibieron quimioterapia. NR4A2 podría servir como factor pronóstico y predictivo y objetivo terapéutico para pacientes con GC (Yifang Han et al; 2013). NR4A2 surge como un factor nuclear importante que vincula la inflamación gastrointestinal y el cáncer, especialmente el CCR, y debería servir como objetivo terapéutico candidato para el cáncer gastrointestinal relacionado con la inflamación (Yi-Fang Han et al; 2012)

Continúa en la siguiente página

Cuadro 14 – continuación de la página anterior

Factor de transcripción	Descripción
NRF1	Vincula la modulación transcripcional de genes metabólicos clave con el crecimiento y desarrollo celular. Implicado en el control de genes nucleares necesarios para la respiración, la biosíntesis del grupo hemo ¹¹ y la transcripción y replicación de ADN mitocondrial
OTX1	Sobre-expresión de tejido (sangre, retina, glándula salival, piel). Desempeña un papel en el desarrollo del cerebro y los órganos sensoriales (Uniprot)
PAX2	Sobre-expresión de tejido (conducto deferente, epidídimo, riñón). TF que puede tener un papel en la diferenciación de células renales (Moumita Barua et al; 2014). Tiene un papel crítico en el desarrollo del tracto urogenital, los ojos y el SNC (Uniprot)
PAX4	Juega un papel crítico durante el crecimiento del cáncer. Actividad viral: el dominio de activación parece responder a E1A mientras que el dominio de represión es activo en general (Fujitani Y et al; 1999). Uno de los TF principales implicados en genes responsables de la expresión diferencial en CCR temprano (Manika Sehgal et al;2015).
PAX5	Sobre-expresión de tejido (sangre, tejido linfoide). Proto-oncogen. Puede desempeñar un papel importante en la diferenciación de las células B, así como en el desarrollo neuronal y la espermatogénesis. Contribuye a la patogenicidad de linfomas de células grandes (Uniprot).
PAX6	Sobre-expresión de tejido (cerebro, retina). TF con funciones importantes en el desarrollo del ojo, nariz, sistema nervioso central y páncreas. Necesario para la diferenciación de las células alfa de los islotes pancreáticos. Compite con PAX4 en la unión a un elemento común en los promotores de glucagón, insulina y somatostatina. Regula la especificación de los subtipos de neuronas ventrales estableciendo los dominios progenitores correctos (Uniprot).
PAX8	Sobre-expresión de tejido(riñón, glándula tiroideas). PAX8 es importante para el desarrollo de varios tipos de tumores. Sin expresión en CCR según Gepia.
PAX9	Sobre-expresión de tejido (esófago, tejido linfoide, glándula paratiroides, glándula salival, lengua). TF requerido para el desarrollo normal del timo, las glándulas paratiroides, los cuerpos ultimobranquiales, los dientes, los elementos esqueléticos del cráneo y la laringe, así como las extremidades distales (Uniprot)
PBX1	Probablemente en complejo con MEIS2, está involucrado en la regulación transcripcional por KLF4. Actúa como un activador transcripcional de NKX2-5 y un represor transcripcional de CDKN2B. Muy poca expresión en CCR (Gepia). El TF de homeodominio PBX1 puede convertirse en la oncoproteína quimérica E2A-PBX1 (Chiou-Hong Lin et al; 2019).

Continúa en la siguiente página

¹¹La síntesis de hemo es una vía bioquímica que requiere varios pasos, sustratos y enzimas. Una deficiencia en una enzima o sustrato conduce a la acumulación de intermedios de síntesis de hemo en sangre, tejidos y orina, lo que lleva a un resultado clínicamente significativo de un grupo de trastornos llamados porfirias.[119]

Cuadro 14 – continuación de la página anterior

Factor de transcripción	Descripción
PGR	Sobre-expresión de tejido (uterino, endometrio, músculo liso). Las hormonas esteroides y sus receptores están involucrados en la regulación de la expresión del gen eucariota y afectan la proliferación y diferenciación celular en los tejidos diana. Dependiendo de la isoforma, el receptor de progesterona funciona como activador o represor transcripcional (Uniprot). Las tasas de incidencia y los riesgos relativos para el CCR son más altos en hombres que en mujeres. Los esteroides sexuales pueden desempeñar un papel en esta diferencia asociada al género en el riesgo de CCR (J Sainz et al; 2011).
PLAG1	TF cuya activación da como resultado la sobre-expresión de genes diana, como IGFII, que conduce a una proliferación celular descontrolada: cuando se sobreexpresa en células cultivadas, se observa una mayor tasa de proliferación y transformación (THPA). El gen PLAG1 es mejor conocido como un oncogén asociado con ciertos tipos de cáncer, involucrado en la proliferación celular mediante la regulación directa de una amplia gama de genes diana, incluidos varios factores de crecimiento. Es probable que esto sea un modo de acción central para PLAG1 tanto en el desarrollo embrionario como en el cáncer (Almas R Juma et al; 2016)
POU1F1	Sobre-expresión de tejido (glándula pituitaria). TF involucrado en la especificación de los fenotipos lactotrofo, somatotrofo y tirotrofo en la hipófisis anterior en desarrollo. Específicamente se une a la secuencia consenso 5'- TAAAT-3 '. Activa los genes de la hormona del crecimiento y la prolactina (THPA)
POU2F1	Sobre-expresión de tejido (cerebro). Sinónimo: oct1. Interactúa con NF-YA. El dominio POU Oct-1 estimula la replicación del ADN de adenovirus mediante una interacción directa entre el precursor viral proteína-ADN terminal, complejo de polimerasa y el homeodominio POU (Coenjaerts, et al; 1994). En un modelo químico de CCR, la pérdida de Oct1 en el colon restringió severamente la tumorigenicidad. (Oct1 / Pou2f1 se requiere selectivamente para la regeneración del colon y regula la malignidad del colon (Karina Vázquez-Arreguín et al; 2019)
POU2F2	Sobre-expresión de tejido (sangre, cerebro, intestino, tejido linfoide). Regula la transcripción en varios tejidos además de activar la expresión del gen de inmunoglobulina. Modula la transactivación de la transcripción por NR3C1, AR y PGR (Uniprot). Según Gepia, muy poca expresión en CCR.
POU3F1	Sobre-expresión de tejido (cerebro, piel). Participa en la neurogénesis y la embiogénesis temprana (Uniprot).
POU3F2	Sobre-expresión de tejido (cerebro). TF que juega un papel clave en la diferenciación neuronal. Se une preferentemente a la secuencia de reconocimiento que consta de dos medios sitios distintos, ("GCAT") y ("TAAT"), separados por una región espaciadora no conservada de 0, 2 o 3 nucleótidos. Actúa aguas abajo de ASCL1, accediendo a la cromatina que ha sido abierta por ASCL1, y promueve la transcripción de genes neuronales (Uniprot).

Continúa en la siguiente página

Cuadro 14 – continuación de la página anterior

Factor de transcripción	Descripción
POU4F2	Sobre-expresión de tejido(cerebro, testículo). Es un TF de unión al ADN específico del tejido implicado en el desarrollo y la diferenciación de las células objetivo. Funciona como activador o represor que modula la tasa de transcripción del gen objetivo a través de la enzima ARN polimerasa II de una manera dependiente del promotor (Ling Zhang et al; 2013). Desempeña un papel en la apoptosis. En cooperación con TP53 potencia la activación transcripcional de la actividad del promotor BAX aumentando la apoptosis de las células neuronales. Regula negativamente la actividad del promotor BAX en ausencia de TP53. Actúa como un coactivador transcripcional a través de su interacción con el TF ESR1 al mejorar su efecto sobre el promotor que contiene el elemento de respuesta al estrógeno (ERE). Antagoniza la actividad estimuladora transcripcional de POU4F1 evitando su unión a un factor de octamero (Uniprot)
PPARA	Los PPAR afectan la expresión de genes diana implicados en la proliferación celular, la diferenciación celular y en las respuestas inmunes e inflamatorias (THPA)
PPARG	Sobre-expresión de tejido(tejido adiposo, mama). Actúa como un regulador crítico de la homeostasis intestinal al suprimir las respuestas proinflamatorias mediadas por NF-kB (Uniprot). Sobreexpresado en CCR (Gepia).
PTF1A	Sobre-expresión de tejido(páncreas). TF implicado en la determinación del destino celular en varios órganos. Se une a la secuencia consenso de E-box 5'-CANNTG-3'. Desempeña un papel en el desarrollo y diferenciación temprana y tardía del páncreas. Importante para determinar si las células asignadas a los brotes pancreáticos continúan hacia la organogénesis pancreática o vuelven a los destinos duodenales. Puede estar involucrado en el mantenimiento de la expresión génica específica del páncreas exocrino, incluyendo ELA1 y amilasa. Requerido para la formación de células pancreáticas acinares y ductales. Desempeña un papel importante en el desarrollo cerebeloso (Uniprot).
RARA	Los heterodímeros RXR / RAR se unen a los elementos de respuesta al ácido retinoico (RARE) compuestos por sitios en tándem 5'-AGGTCA-3' conocidos como DR1-DR5. En ausencia de ligando, los heterodímeros RXR-RAR se asocian con un complejo multiproteico que contiene correpresor ¹² de transcripción que inducen desacetilación de histonas, condensación de cromatina y supresión transcripcional (Harish Srinivas et al; 2006). En asociación con los correpresor HDAC3, HDAC5 y HDAC7, desempeña un papel en la represión de microRNA-10a y, por lo tanto, promueve la respuesta inflamatoria (Ding-Yu Lee et al; 2017)

Continúa en la siguiente página

¹²Un correpresor es una proteína que reduce la expresión génica mediante su unión a un activador o un factor de transcripción que contiene un dominio de unión al ADN. El correpresor es incapaz de unir ADN por sí mismo [121]

Cuadro 14 – continuación de la página anterior

Factor de transcripción	Descripción
RARβ	Sobre-expresión de tejido (retina). Los heterodímeros RXR / RAR se unen a los elementos de respuesta de ácido retinoico (RARE) El heterodímero RXRA / RARβ puede actuar como un represor en el elemento DR1 y como un activador en el elemento DR5. Se une al ácido retinoico, la forma biológicamente activa de la vitamina A que media la señalización celular en la morfogénesis embrionaria, el crecimiento celular y la diferenciación (RefSeq, mar 2014).
RARG	Sobre-expresión de tejido (esófago, piel). Receptor para el ácido retinoico. Los receptores de ácido retinoico se unen como heterodímeros a sus elementos de respuesta diana en respuesta a sus ligandos, ácido retinoico todo-trans o 9-cis, y regulan la expresión génica en varios procesos biológicos (Uniprot). Muy poca expresión en CCR (Gepia). Desregulación de RARG en la leucemia mieloide aguda (Maria Rosa Conserva et al; 2019)
RB1	Supresor de tumor. Regulador clave de entrada en la división celular que actúa como un supresor tumoral. Promueve la transición G0-G1 cuando se fosforila por CDK3 / ciclina-C. Actúa como un represor de la transcripción de genes diana E2F1. La forma activa y menos fosforilada de RB1 interactúa con E2F1 y reprime su actividad de transcripción, lo que lleva a la detención del ciclo celular (Uniprot, THPA). Presente en CCR (Gepia).
RBPJ	Regulador transcripcional con un papel central en la señalización de Notch, una vía de señalización involucrada en la comunicación célula-célula que regula un amplio espectro de determinaciones del destino celular. Actúa como un represor transcripcional cuando no está asociado con las proteínas Notch. Cuando se asocia con algún producto NICD de proteínas Notch (dominio intracelular de Notch), actúa como un activador transcripcional que activa la transcripción de genes diana de Notch.
RELA	NF-κB es un TF pleiotrópico presente en casi todos los tipos de células y es el punto final de una serie de eventos de transducción de señales que se inician por una amplia gama de estímulos relacionados con muchos procesos biológicos como inflamación, inmunidad, diferenciación, células crecimiento, tumorigénesis y apoptosis. NF-κB es un complejo homo o heterodimérico formado por las proteínas RELA/p65, RELB, NFκB1/p105, NFκB1/p50, REL y NFκB2/p52. El RELA-NFκB1 heterodimérico NF-κB funciona como activadores transcripcionales (Uniprot). La vía RelA / NF-κB se expresa constitutivamente en CCR. RelA / NF-κB podría desempeñar un papel importante en la tumorigénesis colorrectal (Ram Rattan Negi et al; 2018)
REST	Representante clave de la expresión génica en hipoxia; reprime los genes en la hipoxia mediante la unión directa a un sitio RE1 / NRSE en sus regiones promotoras (Miguel A S Cavadas et al; 2016). Dependiendo del contexto celular, este gen puede actuar como un oncogén o un supresor tumoral

Continúa en la siguiente página

Cuadro 14 – continuación de la página anterior

Factor de transcripción	Descripción
RORA	Sobre-expresión de tejido(piel). Representa un receptor de ácido retinoico nuclear. La proteína codificada, el receptor de ácido retinoico alfa, regula la transcripción de una manera dependiente de ligando. Este gen ha sido implicado en la regulación del desarrollo, diferenciación, apoptosis, granuloposis y transcripción de genes de reloj (RefSeq, septiembre de 2010)
RREB1	Se ha demostrado que la falta de expresión de miR-143 y miR-145 es una característica frecuente de los tumores colorrectales. REB1 se sobreexpresa en tumores de adenocarcinoma colorrectal y líneas celulares, y la expresión de la transcripción primaria miR-143/145 está inversamente relacionada con la expresión de RREB1. En las líneas celulares de CCR, el grupo miR-143/145 es reprimido por RREB1 aguas abajo de KRAS constitutivamente activo (O A Kent et al; 2013)
RXRA	Desempeña un papel en la atenuación del sistema inmune innato en respuesta a infecciones virales, posiblemente regulando negativamente la transcripción de genes antivirales como los genes IFN tipo I (Ma F, Liu SY et al; 2014). Participa en el proceso biológico del huésped interacción del virus, transcripción y regulación de la transcripción (THPA).
RXRB	Sobre-expresión de tejido(retina). Los heterodímeros RXR / RAR se unen a los elementos de respuesta de ácido retinoico (RARE). El heterodímero RXRA / RARB puede actuar como un represor en el elemento DR1 y como un activador en el elemento DR5. Se une al ácido retinoico, la forma biológicamente activa de la vitamina A que media la señalización celular en la morfogénesis embrionaria, el crecimiento celular y la diferenciación (RefSeq, mar 2014).
RXRG	Sobre-expresión de tejido(esófago, piel). Receptor para el ácido retinoico. Los receptores de ácido retinoico se unen como heterodímeros a sus elementos de respuesta diana en respuesta a sus ligandos, y regulan la expresión génica en diversos procesos biológicos. Los heterodímeros RAR / RXR se unen a los elementos de respuesta al ácido retinoico (RARE) conocidos como DR1-DR5. En ausencia de ligando, actúa principalmente como un activador de la expresión génica (THPA)
SF1	Proceso biológico: procesamiento de ARNm, empalme de ARNm, transcripción, regulación de la transcripción (Uniprot)
SIX3	Sobre-expresión de tejido (cerebro, glándula paratiroides, glándula pituitaria, retina). Regulador transcripcional que puede actuar como un represor y activador transcripcional al unir una secuencia de reconocimiento de núcleo de homeodominio ATTA en estos genes diana (Uniprot). Participa en muchos procesos biológicos. No presente en CCR (Gepia)
Continúa en la siguiente página	

Cuadro 14 – continuación de la página anterior

Factor de transcripción	Descripción
SMAD1	SMAD1, STAT1 y Foxk1 pueden estar involucrados en la regulación transcripcional de miR-32, y su regulación al alza en los tejidos de CCR se informó anteriormente, donde aumentó la proliferación, migración e invasión, y redujo la apoptosis de las células de CCR al inhibir la expresión de fosfatasa y homólogo de tensina (PTEN) (Weiyun Wu et al; 2019). La expresión SMAD4 es un marcador pronóstico en CCR. Las expresiones nucleares de pSMAD1,5,8 y pSMAD2,3 no son marcadores pronósticos útiles en CCR (Philip W Voorneveld et al; 2013)
SNAI1	Participa en la inducción de la EMT, formación y mantenimiento del mesodermo embrionario, detención del crecimiento, supervivencia y migración celular. Marcador pronóstico en CCR (desfavorable). Durante la EMT, participa con LOXL2 en la regulación negativa de la transcripción de heterocromatina pericentromérica. SNAI1 recluta LOXL2 en regiones pericentroméricas para oxidar la histona H3 y reprimir la transcripción que conduce a la liberación del componente de heterocromatina CBX5 / HP1A, lo que permite la reorganización de la cromatina y la adquisición de rasgos mesenquimales (Uniprot)
SOX2	Sobre-expresión de tejido(cerebro). La sobreexpresión y la amplificación génica de Sox2 se han asociado con la agresión tumoral y la metástasis en varios tipos de cáncer, incluidos los de mama, próstata, pulmón, ovario y colon (Sameer Chaudhary et al; 2019). SOX2 mejora la metástasis de las células de cáncer de mama y próstata al promover la EMT a través de WNT / B-catenina (Xuefei Li et al; 2013)
SOX5	Sobre-expresión de tejido(testículos) El ARN largo no codificante ¹³ lnc-sox5 modula la tumorigénesis de CCR al desequilibrar el microambiente tumoral (Wu K, Zhao Z, et al, 2017)
SOX9	Sobre-expresión de tejido(glándula salival). La fuerte expresión de SOX9 es un indicador independiente para un pronóstico adverso en el CCR (Bingjian Lü, MD et al.; 2008). SOX9 interactúa con diferentes TFs y presenta varias características pro-oncogénicas, incluida la promoción de la proliferación, la inhibición de la senescencia y la transformación neoplásica en colaboración con otros oncogenes. Interactúa con: TCF4-regula positivamente el CCR SOX9; La activación de PPAR-influye influye de manera desigual en la expresión de SOX9 y catenina y la localización subcelular, lo que sugiere un papel mecanicista disponible en la carcinogénesis de colon; NF-Y-SOX9 es necesario para la función de la expresión de inactivación de NF-Y de ciclina B1, ciclina B2, quinasa dependiente de ciclina 1; FOXK2-Es un objetivo SOX9 y promueve la proliferación (MaribelAguilar-Medin; 2019). La regulación positiva de SOX9 es común en el adenoma colorrectal y el cáncer.
Continúa en la siguiente página	

¹³Los ARN largos no codificantes (LncRNA) se han considerado recientemente como reguladores sistémicos en múltiples procesos biológicos, incluida la tumorigénesis[120]

Cuadro 14 – continuación de la página anterior

Factor de transcripción	Descripción
SP1	Gen relacionado con el cáncer. Se une con alta afinidad a factores ricos en GC y regula la expresión de genes involucrados en procesos como el crecimiento celular, la apoptosis, la diferenciación y las respuestas inmunes. Mantiene la actividad de la telomerasa en las células cancerosas al inducir la expresión del gen TERT y TERC (THPA). El efecto de los sitios Sp1 depende de la distancia a la caja TATA (Segal R et al; 1991). Cooperación altamente específica con NF-kB (Majello B et al; 1994). Interactúa física y funcionalmente con c-Jun y puede reclutarlo para promotores sin unión de ADN por c-Jun (Kardassis D et al; 1999)
SP2	Miembro de la subfamilia Sp de factores de transcripción Sp/XKLF. Las proteínas de la familia Sp son proteínas de unión a ADN específicas de secuencia caracterizadas por un dominio de transactivación amino terminal y tres factores de dedos de zinc carboxi terminales [RefSeq, jul 2008]
SP3	Factor transcripcional que puede actuar como un activador o represor dependiendo de la isoforma y/o modificaciones postraduccionales. Se une a elementos promotores de cajas GT y GC. Compite con SP1 por los promotores de GC-box. Activador débil de la transcripción, pero puede activar varios genes involucrados en diferentes procesos, como la regulación del ciclo celular (THPA). La interacción cooperativa entre NF-kB y Sp1 es altamente específica, ya que ni Sp3 ni Sp4 son capaces de cooperar con NF-kB (Majello B et al; 1994)
SP8	Sobre-expresión de tejido(próstata). TF que juega un papel clave en el desarrollo de las extremidades (THPA)
SPDEF	Sobre-expresión de tejido(próstata, glándula salival). Participa en la regulación de la glándula prostática y/o el desarrollo del cáncer de próstata (Uniprot)
SREBF1	Proceso biológico: metabolismo del colesterol, metabolismo de lípidos, metabolismo de esteroides, metabolismo de esteroides, transcripción, regulación de la transcripción (Uniprot). Las proteínas de unión a elementos reguladores de esteroides (SREBP) son una familia de TFs que regulan la homeostasis de los lípidos al controlar la expresión de una gama de enzimas requeridas para el colesterol endógeno, el ácido graso, la síntesis de triacilglicerol y fosfolípidos (Delphine Eberlé et al; 2004)
SREBF2	Activador transcripcional requerido para la homeostasis lipídica. Regula la transcripción del gen del receptor de LDL, así como el colesterol y, en menor grado, la vía de síntesis de ácidos grasos (THPA). Proceso biológico: metabolismo del colesterol, metabolismo de los lípidos, metabolismo de los esteroides, metabolismo del esterol, transcripción, regulación de la transcripción (Uniprot)
STAT1	STAT1 participa en la respuesta antiviral y bacteriana, inhibición del crecimiento, estimulación de la apoptosis y supresión del crecimiento tumoral. STAT1 desempeña un papel supresor tumoral en las células cancerosas. Se encontró pérdida de activación y / o expresión de STAT1 en células malignas derivadas de varios tipos histológicos de tumor (Yannick Verhoevena et al; 2019). Papel general en el inicio de la señalización en múltiples tipos de células y función en la activación y diferenciación de las células T (Mahmood Y. Bilal; 2018)

Continúa en la siguiente página

Cuadro 14 – continuación de la página anterior

Factor de transcripción	Descripción
STAT3	La activación STAT3 persistente se dedica a la mitogénesis, supervivencia, antiapoptosis, metástasis y oncogénesis. STAT3 es un jugador importante en el desarrollo y / o progresión del cáncer. Se ha estimado que la actividad STAT3 elevada de forma aberrante ocurre en más del 70 % de los cánceres humanos. La activación de STAT3 induce proteínas que impulsan la proliferación tumoral (como la ciclina D1 y c-Myc) (Yannick Verhoevena et al; 2019)
STAT4	Sobre-expresión de tejido(sangre, testículos). El silenciamiento del gen STAT4 suprimió la proliferación celular y la invasión de células CCR. El aumento de la expresión de STAT4 se correlaciona positivamente con la profundidad de la invasión en pacientes con CCR, y la inhibición de la expresión de STAT4 reprime el crecimiento y la invasión de las células de CCR; STAT4 puede ser un objetivo terapéutico prometedor para el tratamiento de CCR (Yannick Verhoevena et al; 2019)
STAT5A	La expresión de STAT5A juega un papel carcinogénico en la tumorigénesis de las células de CG a través de la reprogramación del metabolismo intracelular de ácidos grasos, que establece un nuevo mecanismo para la tumorigénesis de las células de CG (S-R Dong et al; 2019). La inhibición de STAT5a promovió significativamente la apoptosis de células CCR. La inhibición de STAT5a puede servir como un nuevo objetivo potencial para el tratamiento de CCR (Xuan Hong et al; 2012). Se descubrió que tanto STAT5a como STAT5b estaban involucrados en el crecimiento celular, la progresión del ciclo celular y la apoptosis de las células CCR, y ejercieron sus efectos a través de la regulación de los objetivos posteriores de los genes STAT. Sin embargo, STAT5b influyó en la apoptosis de células CCR más que STAT5a (Wan Du et al; 2012)
STAT6	STAT6 desempeña un papel destacado en la inmunidad adaptativa mediante la transducción de señales de citocinas extracelulares (Huihui Chen et al; 2011)
TAL1	Sobre-expresión de tejido (médula ósea). Implicado en la génesis de neoplasias hemopoyéticas. Puede desempeñar un papel importante en la diferenciación hemopoyética. Sirve como un regulador positivo de la diferenciación eritroide (Uniprot). Implicado en la leucemia mieloide (Yifan Wu et al; 2019) (E R Vagapova et al; 2018)
TBP	Proceso biológico: interacción virus-huésped, transcripción, regulación de la transcripción (Uniprot). Se une al promotor central para colocar la polimerasa adecuadamente, sirve como andamio para el ensamblaje del resto del complejo de transcripción y actúa como un canal para las señales reguladoras. TFIID se compone de la proteína de unión a TATA (TBP) y un grupo de proteínas conservadas evolutivamente conocidas como factores asociados a TBP (THPA)
TCF3	La señalización desregulada de Wnt / B-catenina promueve el CCR al activar la expresión del protooncogen c-MYC. TCF3, juega un papel importante en la regulación de la expresión de MYC en CCR (Meera Sha et al; 2015)

Continúa en la siguiente página

Cuadro 14 – continuación de la página anterior

Factor de transcripción	Descripción
TCF4	Sobre-expresión de tejido(sangre). TF que se une al potenciador de inmunoglobulina Mu-E5 / KE5-factor. Participa en el inicio de la diferenciación neuronal (THPA). Proceso biológico: diferenciación, neurogénesis, transcripción, regulación de la transcripción (Uniprot)
TEAD2	TF que juega un papel clave en la vía de señalización del Hippo, una vía implicada en el control del tamaño de los órganos y la supresión tumoral al restringir la proliferación y promover la apoptosis. Participa en la regulación de la proliferación celular, la migración y la inducción de la EMT (Maren Diepenbruck et al; 2014).
TFAP2A	Sobre-expresión de tejido (placenta, retina). Proteína de unión a ADN específica de secuencia que interactúa con elementos potenciadores virales y celulares inducibles para regular la transcripción de genes seleccionados. Los TFs AP-2 suprimen varios genes, incluidos MCAM/MUC18, CEBPA y MYC (THPA, Lei Yu, Michael J Hitchler et al; 2009). TFAP2A regula la migración de células tumorales y la apoptosis (Francesca Orso et al; 2007). TFAP2A y TFAP2G regulan la progresión tumoral a través de programas genéticos específicos (Francesca Orso et al; 2008). Represor de genes específicos del hígado en células no hepáticas (Y Ren et al; 2001)
TFAP2B	Sobre-expresión de tejido(epidídimo) Proteína de unión a ADN específica de secuencia que interactúa con elementos potenciadores virales y celulares inducibles para regular la transcripción de genes seleccionados. También suprime una serie de genes, incluidos MCAM / MUC18, CEBPA y MYC (THPA).
TFAP2C	Sobre-expresión de tejido(placenta, piel). Proteína de unión a ADN específica de secuencia que interactúa con elementos potenciadores virales y celulares inducibles para regular la transcripción de genes seleccionados. Los factores AP-2 se unen a la secuencia consenso 5'-GCCNNNGGC-3' y activan los genes involucrados en un amplio espectro de funciones biológicas importantes que incluyen el desarrollo adecuado de los ojos, la cara, la pared corporal, las extremidades y el tubo neural. También suprimen una serie de genes, incluidos MCAM/MUC18, CEBPA y MYC (Uniprot)
TFAP4	TF que activa los genes virales y celulares al unirse a la secuencia simétrica de ADN 5'-CAGCTG-3 '(THPA). En los tejidos humanos de CCR, los genes que codifican USP22 y AP4 se sobreexpresan en las lesiones hepáticas metastásicas en comparación con los tejidos de cáncer primario, y su sobreexpresión se asocia significativamente con una pobre supervivencia de los pacientes con CCR. USP22 y AP4 pueden servir como marcadores pronósticos para predecir el riesgo de desarrollar metástasis a distancia en CCR (Yongmin Li et al; 2017). Es un mediador central y coordinador de la progresión del ciclo celular en respuesta a las señales mitogénicas y la activación de c-MYC. Por lo tanto, la inhibición de la función AP4 puede representar un enfoque terapéutico para bloquear la proliferación de células tumorales (Rene Jackstadt, Heiko Hermeking; 2014) Sobreexpresión en CCR (Gepia)

Continúa en la siguiente página

Cuadro 14 – continuación de la página anterior

Factor de transcripción	Descripción
TFDP1	Se une al ADN cooperativamente con los miembros de la familia E2F a través del sitio de reconocimiento E2, que se encuentra en la región promotora de varios genes cuyos productos están involucrados en la regulación del ciclo celular o en la replicación del ADN. El complejo E2F1: DP1 parece mediar tanto la proliferación celular como la apoptosis (Helin K et al; 1991) La mutación TFDP1 indel84 generó un fenotipo de ganancia de función al aumentar la proliferación celular, la migración y la invasión de células CCR (Chen Chen et al; 2014)
THRA, THRB	Sobre-expresión de tejido (cerebro). Receptor de hormona nuclear que puede actuar como represor o activador de la transcripción. Receptor de alta afinidad para las hormonas tiroideas, incluidas la triyodotironina y la tiroxina (THPA). El papel en la generación/progresión del tumor aún no está claro (José M González-Sancho et al; 2003)
TP53	Sobre-expresión de tejido(tejido linfoide). La pérdida de la función p53 contribuye al desarrollo de muchos tipos de cáncer. Actúa como un supresor tumoral en muchos tipos de tumores; induce detención del crecimiento o apoptosis según las circunstancias fisiológicas y el tipo de célula. Participa en la regulación del ciclo celular como un activador trans que actúa para regular negativamente la división celular. Induce la detención del ciclo celular, la apoptosis, la senescencia, la reparación del ADN o los cambios en el metabolismo (Blagih J et al; 2020). Sobre-expresión en CCR por sobre tejido normal (Gepia)
TP73	Sobre-expresión de tejido (cerebro). Supresor de tumor. Participa en la respuesta apoptótica al daño del ADN. Isoformas con dominio de transactivación son pro-apoptóticas, las isoformas que carecen del dominio son anti-apoptóticas y bloquean la función de p53 y transactivan las isoformas de p73 (Uniprot)
TTK	Sobre-expresión de tejido (tejido linfoide, testículo). Fosforila proteínas en serina, treonina y tirosina. Probablemente asociado con la proliferación celular (Uniprot)
VDR	Sobre-expresión de tejido(intestino, glándula paratiroides). VDR: Receptor de vitamina D. Receptor nuclear para calcitriol, la forma activa de vitamina D3 que media la acción de esta vitamina en las células (N Rochel et al; 2000) (Guy Eelen et al; 2005) Sobreexpresada en CCR (Gepia). La disfunción del tejido adiposo se ha relacionado con inflamación de bajo grado, que está relacionada con enfermedades como el cáncer. La expresión de VDR de tejido adiposo alto en CCR modifica la metilación del ADN del tejido adiposo y promueve la inflamación (Daniel Castellano-Castillo et al; 2018).
WT1	Sobre-expresión de tejido(endometrio, trompa de Falopio, ovario). Actúa tanto como un supresor tumoral, así como un oncogen en la formación de tumores, dependiendo del tipo de célula en la que se expresa. Tiene un papel esencial en el desarrollo normal del sistema urogenital y juega un papel en el crecimiento celular, el proceso por el cual las células maduran para realizar funciones específicas y apoptosis (L Yang et al; 2007). Sobre-expresado en cáncer de ovario. (THPA, Gepia)

Continúa en la siguiente página

Cuadro 14 – continuación de la página anterior

Factor de transcripción	Descripción
XBP1	XBP1 se expresa a altos niveles en muestras de tejido metastásico y pobremente diferenciado, mientras que parecía relativamente bajo en muestras de tejido moderadamente y bien diferenciadas en comparación con aquellas en tejido de colon normal. La puntuación de inmunorreactividad XBP1 se asoció positivamente con el aumento de la invasión tumoral. Los estudios <i>in vitro</i> mostraron que la sobreexpresión de XBP1 promovió la invasión de células cancerosas (Nizar M. Mhaidat, et al, 2015)
YY1	Miembro de la familia GLI-Kruppel de proteínas de dedos de zinc que desempeña papeles vitales en muchos procesos biológicos, especialmente la tumorigénesis. YY1 es un inductor de metástasis de cáncer y puede servir como biomarcador clínico y objetivo terapéutico para el tratamiento del cáncer (Wenmeng Wang et al; 2017). YY1 se sobreexpresa en muchos tipos de cáncer y que se demostró que está involucrado en la regulación de la supervivencia celular, la proliferación celular, la invasión celular, la metástasis y la resistencia (Benjamin Bonavida; 2017). Un alto nivel de YY1 y E2F1, o un bajo nivel de miR-526b-3p, predijeron una pobre supervivencia de los pacientes con CCR. Existe una desregulación del eje YY1 / miR-526b-3p / E2F1 en el desarrollo de CCR, lo que implica una nueva vía reguladora para E2F1 como posible objetivo terapéutico en CCR (Zejun Fang et al; 2019). YY1 promueve la proliferación del CCR a través del eje miR-526b-3p / E2F1
ZBTB33	TF que reconoce selectivamente sitios que contienen CpG metilados, así como una diana de ADN específica de secuencia. Los informes cada vez mayores vinculan la sobreexpresión de ZBTB33 y las actividades transcripcionales con el potencial metastásico y el mal pronóstico en el cáncer (Amir Pozner 1 et al; 2016)
ZBTB7A	Dedo de zinc y dominio BTB que contiene 7A - TF que reprime la transcripción de una amplia gama de genes implicados en la proliferación y diferenciación celular. La sobreexpresión de FBI-1 inhibió la activación de AR dependiente de ligando (Jiajun Cui et al; 2011)
ZEB1	Actúa como represor transcripcional. Inhibe la expresión del gen interleucina-2. Reprime el promotor de E-cadherina e induce una EMT reclutando SMARCA4 / BRG1 (Uniprot). Subexpresado en CCR
ZFX	Zfx es un regulador importante del mantenimiento y expansión de células T periféricas (Smith-Raska MR et al; 2018)
ZIC2	Sobre-expresión de tejido (cerebro). Actúa como un activador o represor transcripcional. Desempeña papeles importantes en la etapa temprana de la organogénesis del SNC. Proceso biológico: diferenciación, neurogénesis, transcripción, regulación de la transcripción (Uniprot)
ZIC3	Sobre-expresión de tejido (cerebro). Cáncer enriquecido (cáncer de testículo). Participa en procesos biológicos de diferenciación, neurogénesis, transcripción, regulación de la transcripción (THPA). Las células madre embrionarias (ESC) deben pasar por una serie de estados celulares intermedios antes de diferenciarse terminalmente. ZIC3 es un regulador potencial de la transición del destino celular (Shen-Hsi Yang et al; 2019)

Continúa en la siguiente página

Cuadro 14 – continuación de la página anterior

Factor de transcripción	Descripción
ZIC4	Sobre-expresión de tejido(cerebro). Miembro de la familia ZIC de proteínas de dedo de zinc tipo C2H2: importante durante el desarrollo, y se ha asociado con heterotaxia visceral ligada al cromosoma X y holoprosencefalia tipo 5. Relacionado con el gen que codifica la proteína de dedo de zinc del cerebelo 1 (THPA)
ZNF423	Célula inmune enriquecida (célula NK). Funciona como un TF de unión al ADN mediante el uso de distintos dedos de zinc en diferentes vías de señalización. Por lo tanto, se cree que este gen puede tener múltiples funciones en la transducción de señales durante el desarrollo (THPA)

Cuadro 14: Descripción de los factores de transcripción encontrados. Se describe el rol determinado en CCR (en caso de existir).

13. Similitudes y diferencias entre los factores de transcripción de los 8 promotores

Considerando la descripción realizada de cada proteína seleccionada para realizar el trabajo en la sección de *Introducción*, es de suponer que los factores encontrados para cada promotor difieran. Dado que las funciones de las proteínas son particulares en CCR, encontrar los factores asociados a cada una de ellas es de gran importancia a la hora de comprender las similitudes y diferencias entre los distintos roles. La tabla 15 es una síntesis de los factores encontrados para cada uno de los promotores en función de las variables elegidas para el estudio de promotores híbridos. La clasificación que se presenta en la tabla es mutuamente excluyente, es decir, si el factor esta en la columna de específico de tejido, no es específico de cáncer y vice versa. En el caso de tener ambas características se encuentra en la tercer columna.

Promotor	Específico de cáncer	Específico de tejido	Ambas
A33	KLF4, TP53, FLI1, MYB	-	CDX2
AQP5	STAT3, NFKB1, TP53, STAT1, SOX2, FOXA1, E2F1, NF1, WT1, JUN, ESR1, RELA, NFE2L2, RARG, SMAD1, FOS, EGR1, STAT5A, ESR2, YY1, SP1, CUX1, ETS1, EP300, IKZF1, MYOD1, E2F4, ETV4, IRF1, STAT6, ITGAM	-	VDR, NR112

CDC25A	TP53, E2F1, EZH2, AR, ESR1, EP300, YY1, JUN, E2F4, NFKB1, FOXO3, FOXM1, TFDP1, EGR1, STAT3, PPARG, RELA, XBP1, TAL1, SP1, STAT1, INSM1, IRF1, PBX1, E2F5, FOS, LEF1, FOXO1, MYB, ETV4, WT1, ETS1, RARA, STAT5A, STAT6, TCF3, NF1, ESR2, ELK1, RB1	-	KLF5, HNF4A
CDC25B	TP53, FOS, NFKB1, JUN, ETS1, XBP1, E2F1, E2F4, RELA, PGR, GRB2, SP1, NF1, TP73, EGR1, AHR, TFDP1, ETV5, STAT1, CREB1, EHF, CTCFL, KLF4, RARA, IRF1, ETV1, NFATC2, FOXA1, MEF2A	-	KLF5, HNF1A, ELF3, POU2F2
FSCN1	TP53, NFKB1, GRB2, STAT1, FOS, NF1, YY1, AHR, E2F1, ESR1, ETS1, LEF1, RARA, CREB1, XBP1, WT1, IKZF1, AR, JUN, NFATC2, KLF4, IRF1, ARNT, E2F4, PAX5, SP1, STAT4, RREB1, NR4A2, ETV5, ETV4, CTCFL, HOXA13, TFDP1	HOXB13	HNF4A, KLF5
GLI1	TP53, AR, E2F1, LEF1, STAT1, ETS1, NFKB1, NF1, PAX8, PAX5, EP300, RELA, STAT5A, ARNT, KLF4, NFE2L2, AHR, MECOM, SP1, RARG, LHX1, PGR, FOXO3, EGR1, FOXM1, DDIT3, WT1, E2F4, TCF3, ETV4, MYOD1, TAL1, INSM1, EWSR1, TEAD2, FOXD3, MAFB, MYB, PLAG1, TFDP1, KLF6, IRF1	-	VDR, KLF5, CDX1
NQO1	TP53, NFE2L2, PPARG, FOS, JUN, ESR2, XBP1, AR, PGR, IRF1, FOXA1, KLF4, NR1H2, BACH1, E2F1, STAT1, LEF1, HIC1, RARG, EWSR1, NF1, SP1, RARA, MAF, MYB, PAX5, ETS1, STAT6, GRB2, MECOM, MAZ, STAT4, YY1	-	HNF4A
VWA1	TP53, FOXA1, FOXD3, FOS, STAT4, SOX5, SOX9, XBP1, MNX1	-	HNF4A

Cuadro 15: Factores de transcripción de los ocho promotores en función de las variables *Específico de tejido*, *Específico de cáncer* o *Ambas*.

Se puede ver claras diferencias en dichos resultados. Por un lado, A33 es el único promotor que presenta el factor CDX2. La figura 25 muestra los niveles de expresión de la proteína. En el mapa corporal se observa en rojo que el único tejido tumoral en el cual se sobre-expresa la proteína es en el colon y recto. En el análisis cuantitativo se observa una sobre-expresión de la proteína en COAD (Colon Adenocarcinoma) y READ (Rectum Adenocarcinoma). Por lo tanto, esta especificidad en CCR hace a la especificidad de la proteína A33 en CCR. Los resultados con CDX1, encontrado en GLI1, son muy similares a los presentados en la figura 25.

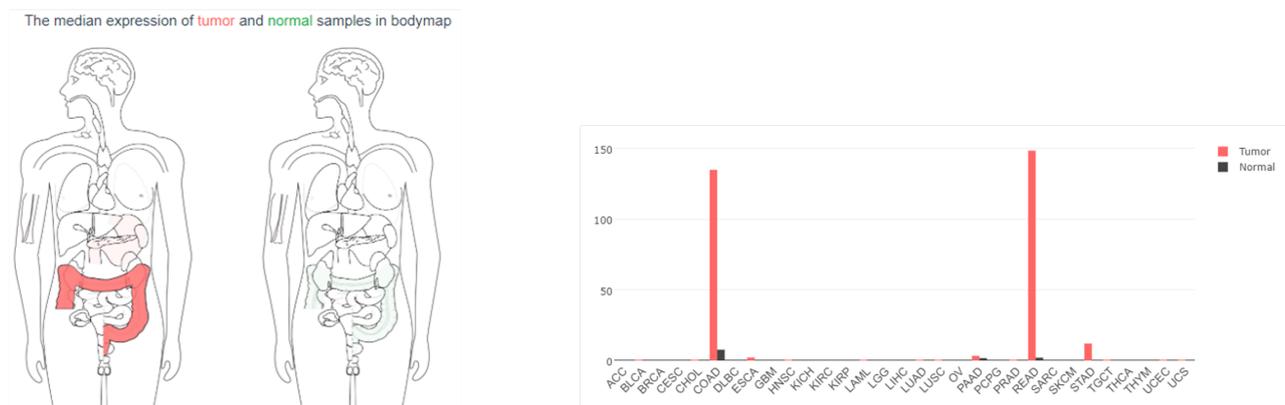


Figura 25: Niveles de expresión de CDX2. A la **izquierda** se observa un mapa corporal, y en rojo se marca los tejidos necróticos donde se ve una sobre-expresión de la proteína. A la **derecha** se observa el análisis cuantitativo. Información obtenida de Gepia.

Con respecto a KLF5 presente en CDC25A, CDC25B, FSCN1 y GLI1, y HNF4A, presentes en CDC25A, FSCN1, NQO1 y VWA2, los niveles de sobre-expresión en CCR no son tan buenos como con CDX2 ya que tal como se refleja en la figura 26 y 27 respectivamente, la sobre-expresión de la proteína se encuentra en varios tejidos tumorales. Sin embargo, se ve una clara sobre-expresión de CCR sobre el tejido normal intestinal tanto en el mapa como en el análisis cuantitativo. En particular, HNF4A tiene sobreexpresión en una menor cantidad de tejidos tumorales comparado con KLF5, lo cual lo hace más específico.

Con respecto a HNF1A, presentado en la figura 28, el mismo tiene un perfil muy similar a HNF4A. HNF1A se encontró en CDC25B.

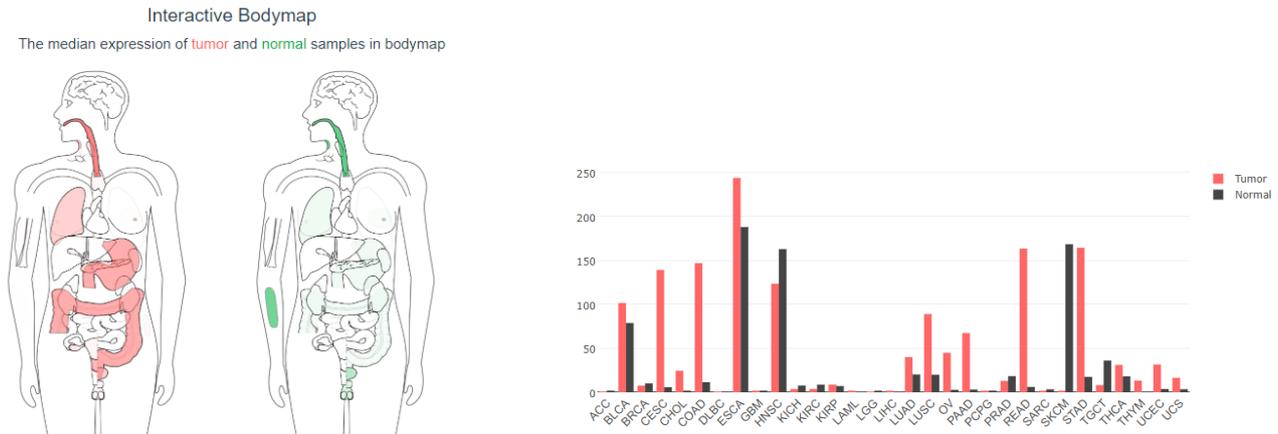


Figura 26: Niveles de expresión de KLF5. A la **izquierda** se observa un mapa corporal, y en rojo se marca los tejidos necróticos donde se ve una sobre-expresión de la proteína. A la **derecha** se observa el análisis cuantitativo. Se deben ver las abreviaciones de tumor COAD y READ. Información obtenida de Gepia.

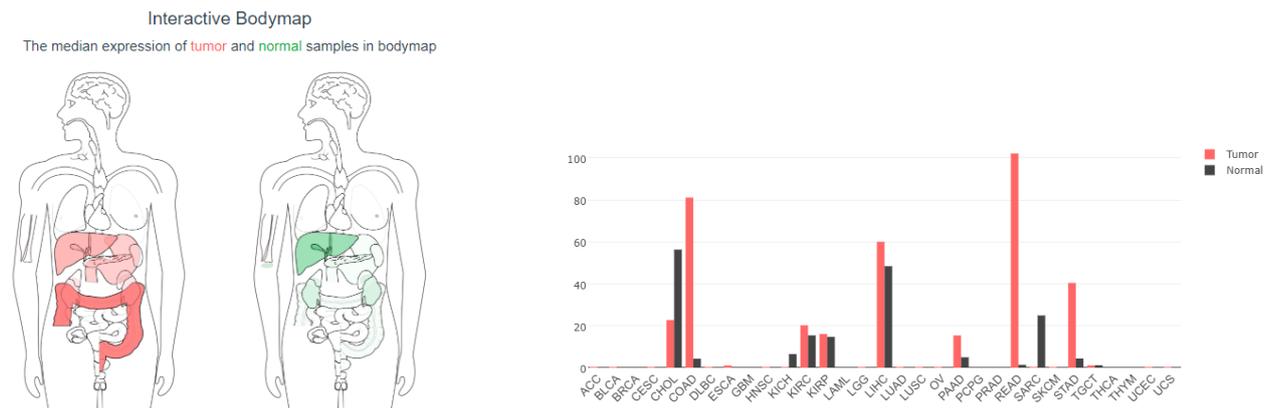


Figura 27: Niveles de expresión de HNF4A. A la **izquierda** se observa un mapa corporal, y en rojo se marca los tejidos necróticos donde se ve una sobre-expresión de la proteína. A la **derecha** se observa el análisis cuantitativo. Información obtenida de Gepia.

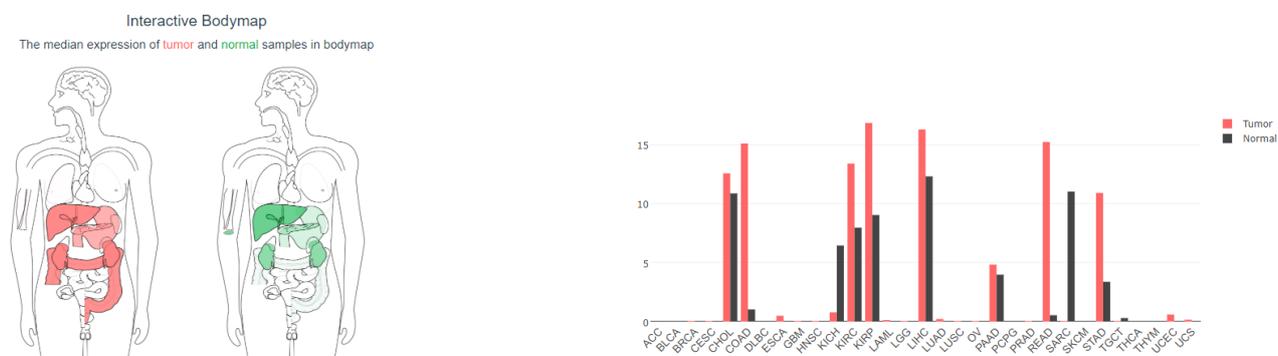


Figura 28: Niveles de expresión de HNF1A. A la **izquierda** se observa un mapa corporal, y en rojo se marca los tejidos necróticos donde se ve una sobre-expresión de la proteína. A la **derecha** se observa el análisis cuantitativo. Información obtenida de Gepia.

Con respecto a VDR presente únicamente en AQP5 y GLI1, la expresión de dicha proteína se encuentra en muchos tejidos tumorales, pero la sobre-expresión por sobre el tejido normal en CCR es muy notoria. La figura 29 refleja estos resultados.

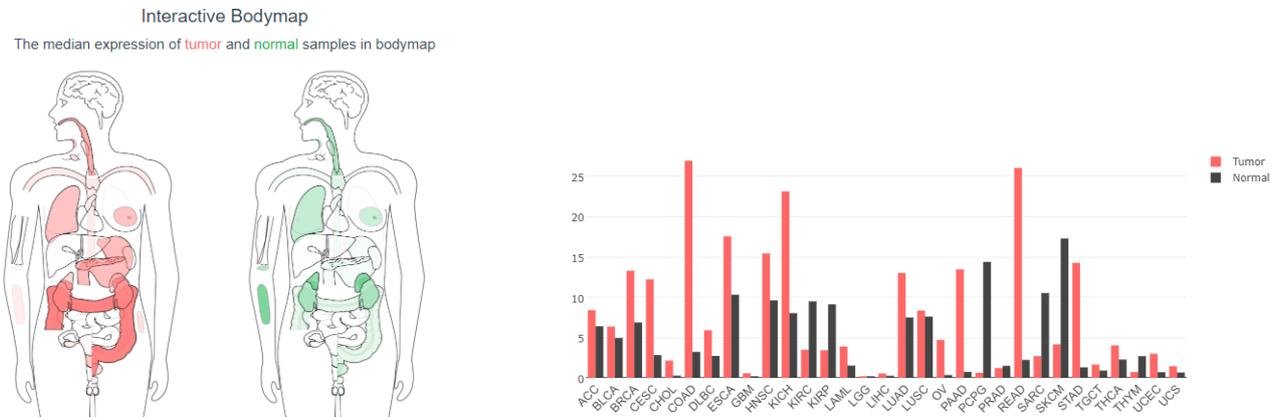


Figura 29: Niveles de expresión de VDR. A la **izquierda** se observa un mapa corporal, y en rojo se marca los tejidos necróticos donde se ve una sobre-expresión de la proteína. A la **derecha** se observa el análisis cuantitativo. Información obtenida de Gepia.

NR1I2 presenta un perfil más exacto a la hora de analizar la sobre-expresión de la proteína, similar al de CDX1 y CDX2. En la figura 30 se observa una clara sobre-expresión en el tejido colorrectal necrótico por sobre el normal. A su vez, la sobre-expresión de la proteína se observa únicamente en CCR y en tejido normal de hígado. NR1I2 se encuentra presente únicamente en AQP5.

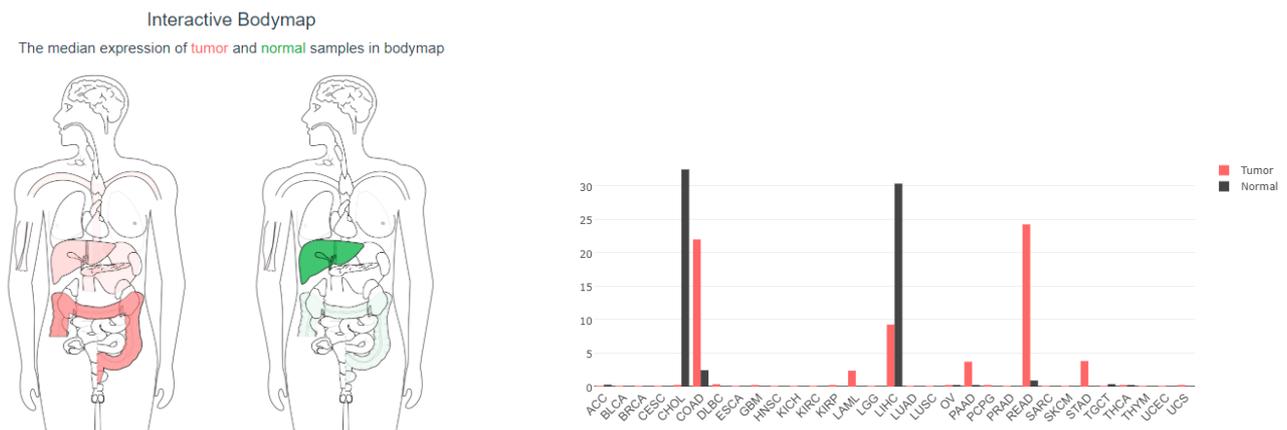


Figura 30: Niveles de expresión de NR1I2. A la **izquierda** se observa un mapa corporal, y en rojo se marca los tejidos necróticos donde se ve una sobre-expresión de la proteína. A la **derecha** se observa el análisis cuantitativo. Información obtenida de Gepia.

El factor ELF3 se encontró únicamente en CDC25B. El mismo tiene un perfil de expresión más amplio, ya que se observa en numerosos tejidos tumorales. Sin embargo, nuevamente se refleja una sobre-expresión en CCR por sobre tejido normal del colon y el recto. La figura 31 refleja estas conclusiones.

cáncer tienen una mayor importancia que el resto ya que hacen a la especificidad y potencia del promotor al CCR y por lo tanto, la elección del fragmento del promotor a utilizar esta guiada por la presencia de dichos factores. A su vez, se priorizó la elección del fragmento según los factores indispensables para la activación del promotor, como FOXM1 en GLI1. En las figuras se esquematiza un mapeo de la secuencia del promotor de cada proteína con los factores asociados a ellos. Dichos factores son los resultados del ranqueo genético de *ToppGene* presentes en la tabla 15. A su vez, teniendo en cuenta la información del rol de cada uno presente en las tablas 13 y 14, se tuvo en cuenta las interacciones entre los mismos.

La elección de cada parte de cada promotor se debe a:

- Presencia de los factores asociados a la especificidad de cáncer y de tejido.
- Presencia de factores identificados como cruciales a la hora de regular la expresión de la proteína. Ejemplo: NFE2L2 en el promotor NQO1, FOMX1 en GLI1.
- Relación entre los TFs. Ejemplo: en NQO1, NFE2L2 con c-Jun se une al ARE y activa la expresión génica.

Es importante mencionar que en las estructuras en las cuales se encuentre marcada la TATA BOX es porque la misma esta verificada por distintos trabajos científicos en dicha posición. Además, en A33, AQP5, CDC25B, FSCN1 y VWA2 se encontraron TFs tales como TBP y TFIID, los cuales forman parte del complejo de pre-iniciación de la ARN polimerasa II en dichas posiciones, corroborando que la estructura de la caja TATA BOX en las secuencias correspondientes. Por otro lado, para el resto de las proteínas, en todos los casos se encontraron factores asociados al complejo pre-iniciación de la ARN-polimerasa II, sin embargo, dichos factores se presentan en varias posiciones y no se encuentran verificadas por papers. En particular, para GLI1, Cheng Zheng Liu et al., 1998, sostiene que la región que rodea los sitios de inicio de la transcripción carece de secuencias de consenso TATA y CCAAT, tiene un alto contenido de GC, incluye una isla CpG y contiene varias cajas de GC. Sin embargo, con esta metodología se encontraron factores CEBPA, CEBPB y CEBPD asociados al promotor. Por lo tanto, para corroborar estos resultados se debe hacer un estudio *in vitro* de PCR para verificar la ausencia/presencia de dichos factores.

Los factores de cada promotor señalados en las figuras a continuación se presentan en las tablas 5 6 7 8 9 10 11 12. La elección de la presentación y los colores en las figuras subsiguientes se debe a la rápida visualización de cuales factores y en que posiciones se encuentran en dos o tres bases de datos. Esto hace que la posición de dicho factor tenga mayor relevancia por sobre otra posición del mismo factor que solo se encuentran en una plataforma. En todos los casos, el inicio de transcripción (+1) se encuentra marcado en verde, en **negro** son los factores encontrados por PROMO, en **rosa** son los factores encontrados por **FIMO**, en **azul** por MATCH y en **blanco** son aquellos encontrados por 2 o 3 plataformas en la misma posición.

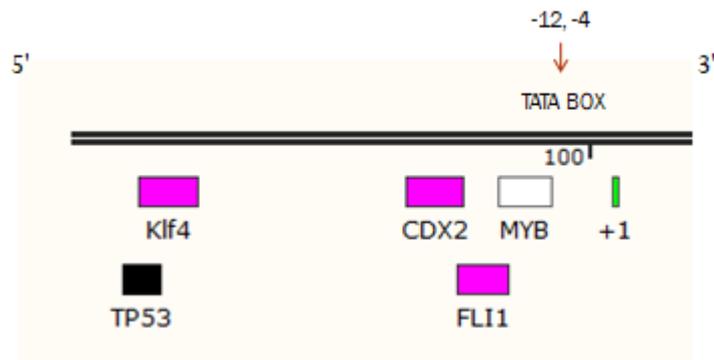


Figura 33: Estructura del promotor A33 con los factores identificados en la tabla 15.

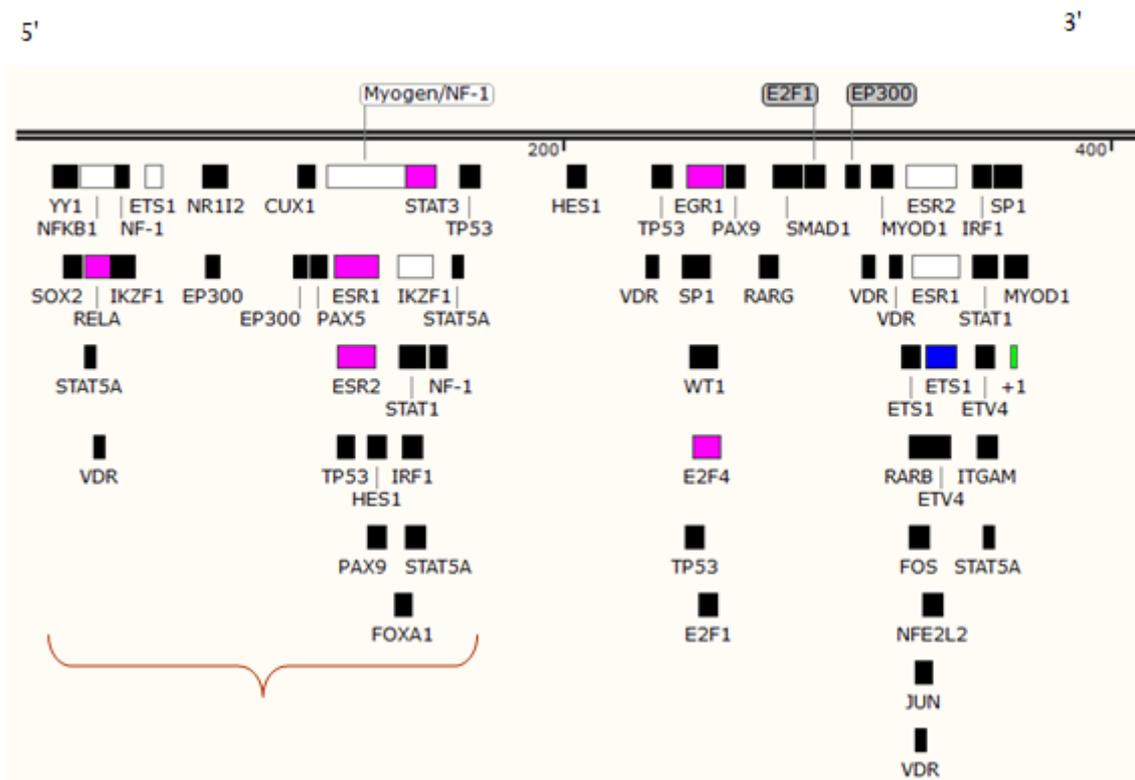


Figura 34: Estructura del promotor AQP5 con los factores identificados en la tabla 15.



Figura 35: Estructura del promotor CDC25A con los factores identificados en la tabla 15.

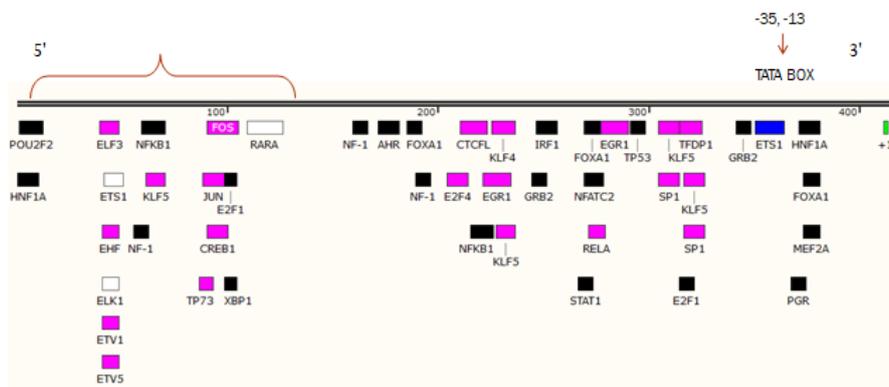


Figura 36: Estructura del promotor CDC25B con los factores identificados en la tabla 15.

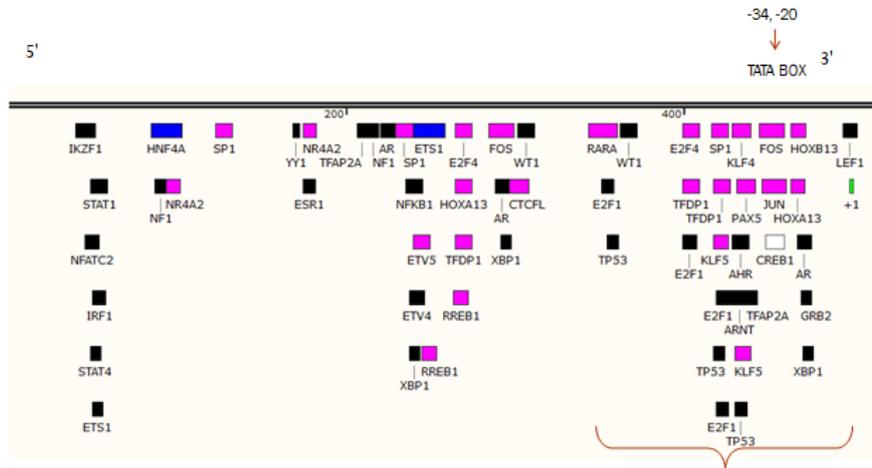


Figura 37: Estructura del promotor FSCN1 con los factores identificados en la tabla 15.

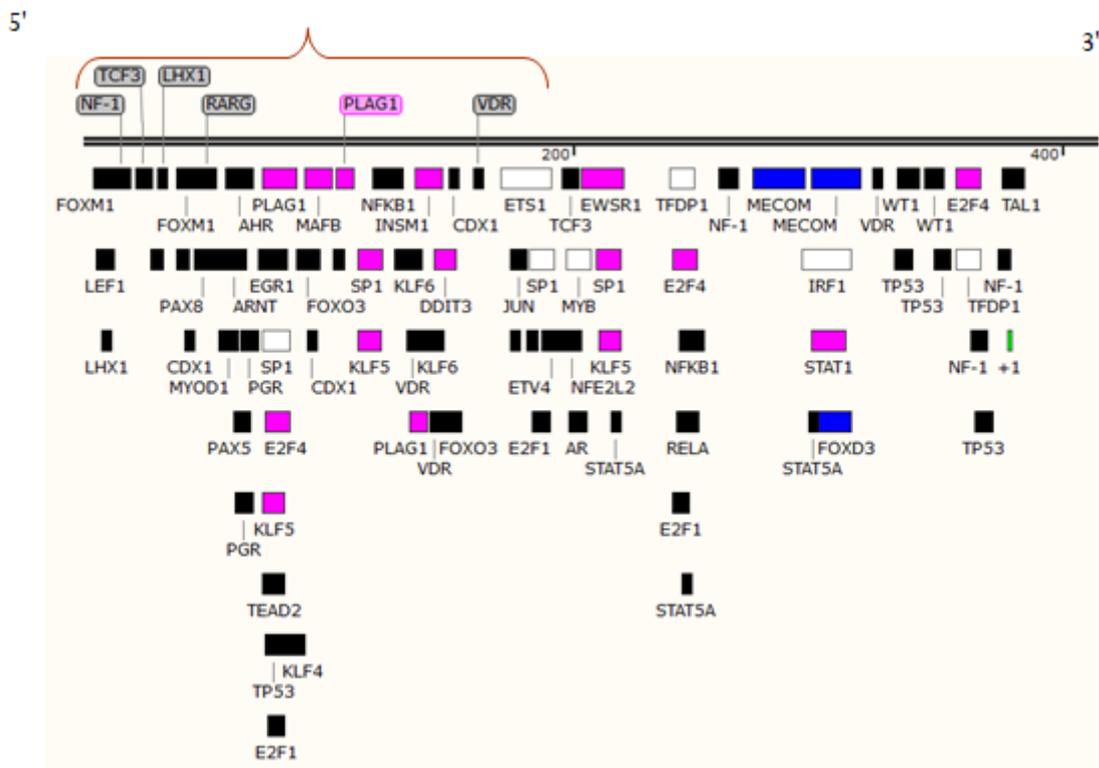


Figura 38: Estructura del promotor GLI1 con los factores identificados en la tabla 15.

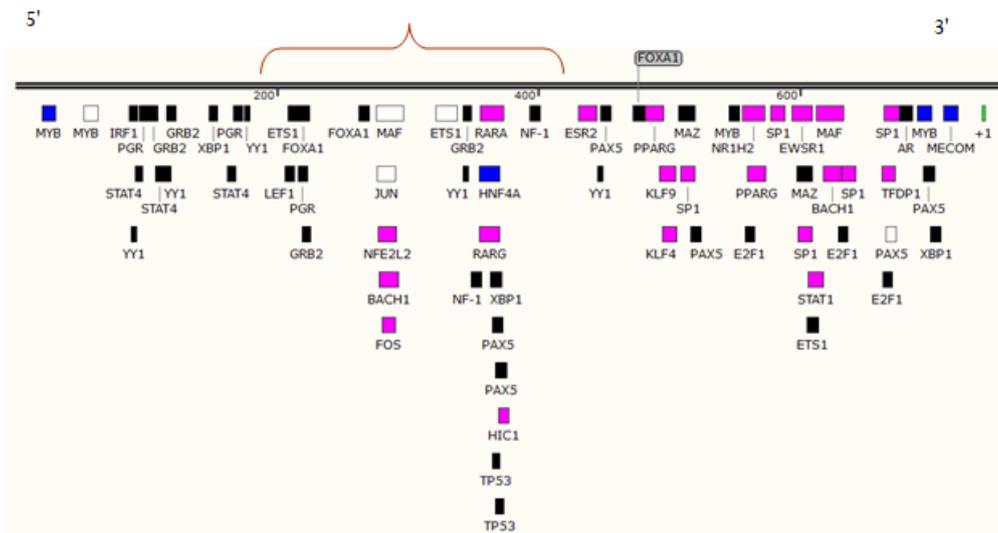


Figura 39: Estructura del promotor NQO1 con los factores identificados en la tabla 15.

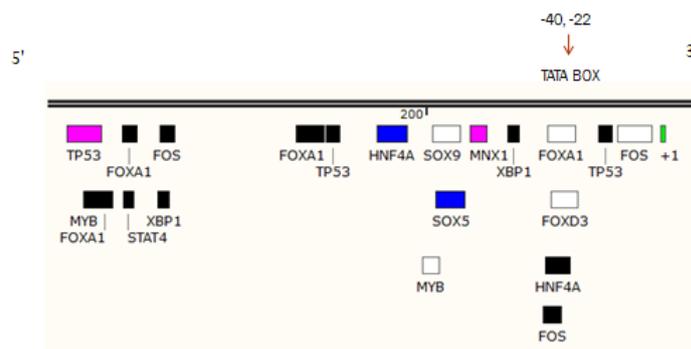


Figura 40: Estructura del promotor VWA2 con los factores identificados en la tabla 15.

Se busca combinar las distintas secciones de los promotores híbridos de manera tal que se complementen los TFs específicos de cáncer y de tumor, teniendo en cuenta cual promotor tiene una caja TATA asociada y cual aún no se sabe con certeza. De esta manera, los futuros modelos se van a armar teniendo en cuenta las características propias de la proteína, la expresión en tumores primarios/metastásicos, los TFs presentes y la presencia confirmada de la caja TATA. La figura 41 presenta un modelo a modo de ejemplo de un posible promotor híbrido que regule la actividad de un AOL-RC para CCR, con los fragmentos indicados en las figuras anteriores.

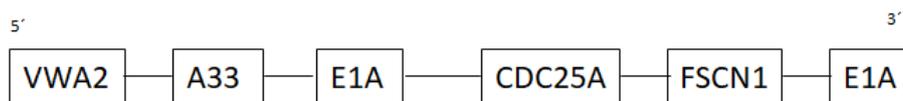


Figura 41: Ejemplo de modelo de promotor híbrido de hasta 4 PETs.

15. Promotores sintéticos

Teniendo en cuenta que en la tabla 15 presenta los resultados de los factores asociados a la especificidad de tejido, cáncer, y a ambas características, a continuación presento los factores asociados a las otras dos variables definidas para el estudio de promotores sintéticos.

Factores asociados a actividad viral: E2F1, EP300, ETV4, EZH2, GRB2, POU1F1, POU2F1, RB1, RELA, RXRA, YY1

Factores asociados a la actividad inmunológica: BACH2, CEBPA, CEBPB, CEBPD, CR1, DDIT3, EGR3, EZH2, FOXP3, HMGB1, IRF1, ITGAM, LEF1, MAF, NFATC4, NR1H4, NR1I2, PPARA, RXRA, STAT6, TAL1, ZFX

Los factores que presentan ambas características, de especificidad de cáncer y de tejido, son CDX1, CDX2, ELF3, HNF1A, HNF4A, KLF5, NR1I2, POU2F2, VDR, y viendo los perfiles de expresión de los mismos, los más fuertes en CCR son CDX1, CDX2, HNF1A, HNF4A, NR1I2, KLF5, ELF3 y VDR. Por otro lado, los factores que presentan ambas características, de actividad inmunológica y actividad viral son EZH2 y RXRA. A su vez, teniendo en cuenta los resultados obtenidos por Manika Sehgal et al., 2015 presentados en la figura 42 se utiliza para complementar con los resultados obtenidos y decidir posibles combinaciones de TF para el armado de los promotores sintéticos.

 Revealed Key Genes in CRC via Novel Devised Parameters

Table 2. Identified major transcription factors in early colorectal cancer progression.

Transcription Factor	Frequency	Importance	JASPAR ID ₁	Class	Family	Pubmed IDs/ Experimental Databases ₂
HNF4	31.80%	0.31802	MA0114.1	Zinc-coordinating	Hormone-nuclear Receptor	19048623, 22731903, 22308320
NR2F1	19.43%	0.50044	MA0017.1	Zinc-coordinating	Hormone-nuclear Receptor	The Human Protein Atlas
DR1	17.31%	0.04112	-	Zinc-coordinating	Hormone-nuclear Receptor	The Human Protein Atlas, 10690519, 19251712
PPARG	14.49%	0.03622	MA0066.1	Zinc-coordinating	Hormone-nuclear Receptor	19186181, 16489531
HNF1	14.13%	0.36064	MA0046.1, MA0153.1	Helix-Turn-Helix	Homeo	12730871, 20096102
HNF4_DR1	13.78%	0.16882	-	Zinc-coordinating	Hormone-nuclear Receptor	22383578, 18180275
PPAR_DR1	13.43%	0.13428	-	Zinc-coordinating	Hormone-nuclear Receptor	11840453, 25961905
HNF4ALPHA	12.01%	0.29848	MA0114.1	Zinc-coordinating	Hormone-nuclear Receptor	25961905, The Human Protein Atlas, 22731903
PAX4	12.01%	0.18322	MA0068.1	Helix-Turn-Helix	Homeo	12970747, The Human Protein Atlas, 19395656
ER	10.60%	0.08216	MA0112.2, MA0258.1	Zinc-coordinating	Hormone-nuclear Receptor	20663982

¹The JASPAR IDs correspond to the transcription factors from JASPAR database
²The Pubmed IDs/ Experimental Databases column contains the information for literature references and databases created on experimentally validated data for their association with colorectal cancer.

Figura 42: Resultados de Manika Sehgal et al, 2015, de 10 TF principales implicados en genes responsables de la expresión diferencial en CCR temprano.

Teniendo en cuenta todo lo estudiado hasta el momento de los promotores sintéticos y los resultados obtenidos estas propuestas incluyen a los factores HNF4A, CDX2, SP1, RXRA/PPARA (ya que es un heterodímero), AP-1, AP-2, NFKB1 y PAX4, con combinaciones aleatorias de los mismos, y la respectiva caja TATA con los TF involucrados. Por lo tanto, las primeras tres combinaciones son HNF4A, CDX2, SP1, RXRA/PPARA, en posiciones aleatorias, en combinación con 3, 4 y 5 de cada TF seleccionado (figura 43). Todas las combinaciones propuestas siguen la línea de los resultados obtenidos por Mohamed et al., y por lo tanto, la distancia fija entre TFs

es de 15pb y la distancia del último TF con la caja TATA es de 55 pb. Es importante mencionar que los ejemplos que se presentan a continuación son a modo de ejemplo con algunos de los factores seleccionados, se pueden realizar diversas combinaciones en función de lo resaltado en 42, 14 y 15. A medida que aumenta la cantidad de TFs, aumenta el tamaño del promotor.

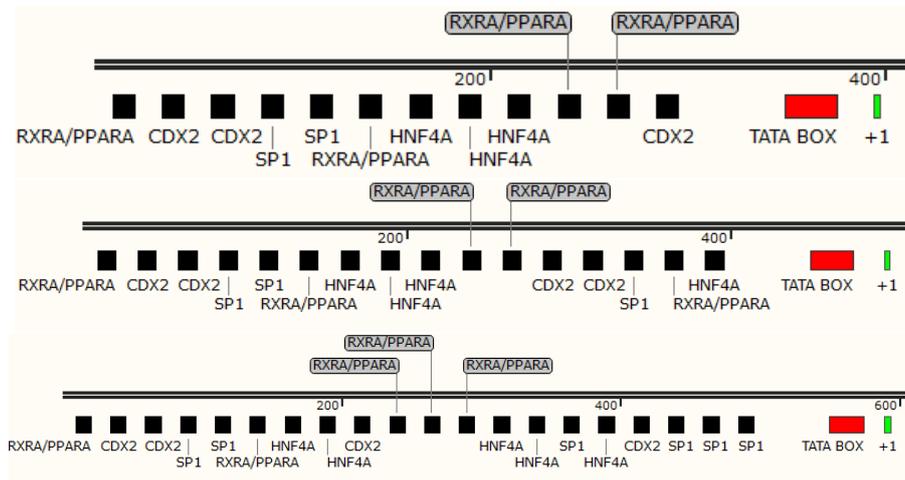


Figura 43: Primeros tres promotores sintéticos propuestos con los factores HNF4A, CDX2, SP1, RXRA/PPARA. Difieren en la cantidad de cada TF seleccionado.

Una segunda propuesta es con más factores, involucrando factores conocidos tales como NFKB, AP2 y AP-1. Teniendo en cuenta la relación específica entre SP1, NFKB y AP1 mencionada en la tabla 14. Una vez mas, los modelos siguen un orden de TFs Manika Sehgal et al, 2015, los cuales se corroboran en las tablas presentadas como resultados aleatorios.

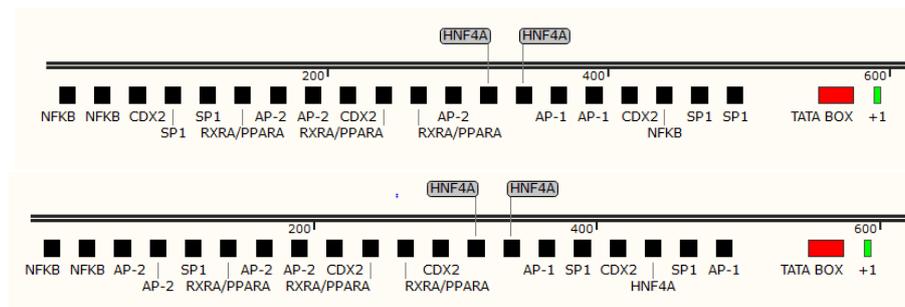


Figura 44: Promotores sintéticos propuestos con los factores HNF4A, CDX2, SP1, RXRA/PPARA, NFKB, AP2 y AP-1. Difieren en la cantidad de cada TF seleccionado.

Otras propuestas son incluyendo al factor PAX4, teniendo en cuenta los resultados obtenidos por Manika Sehgal et al, 2015, los cuales se corroboran en las tablas presentadas como resultados.

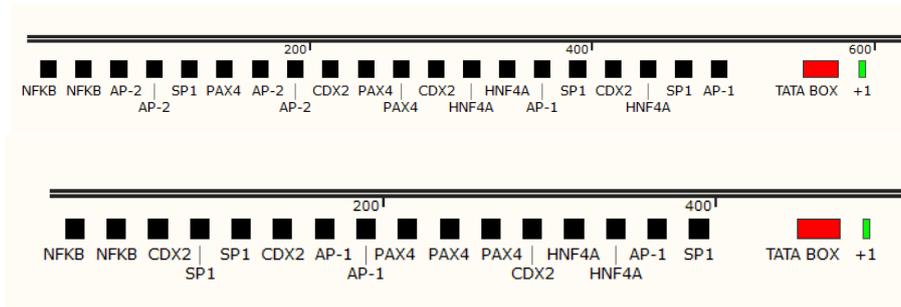


Figura 45: promotor sintético propuesto con los factores HNF4A, CDX2, SP1, PAX4, NFKB, AP2 y AP-1.

Discusión

16. Situación actual de CCR

Hoy en día, las terapias actuales de CCR tienen la desventaja de que no logran atacar todas las poblaciones celulares que conforman un mismo tumor. Según el estadio del mismo, las probabilidades de sobrevivir difieren en función de la terapia implementada. Sin embargo, aquellos clones no sensibles al tratamiento pueden llevar a una nueva proliferación más agresiva. Es decir, la heterogeneidad celular que conforma al tumor es la principal característica de la recidiva de la enfermedad. Incluso las nuevas terapias personalizadas/dirigidas tienen un único blanco terapéutico, y por lo tanto van generando indirectamente que los clones que no sean atacados por la terapia escapen al tratamiento y sean los responsables de la recidiva. Teniendo en cuenta las cifras estimadas de la enfermedad mencionadas en la sección *Introducción*, se vuelve imprescindible la búsqueda de terapias que logren manejar con mayor eficiencia los distintos subclones que conforman el tumor. Una de las terapias que se encuentra en continuo crecimiento consiste en el uso de AOs con una replicación restringida dirigida por PETs dirigiendo la actividad de E1A. La integración de los adenovirus oncolíticos con las terapias convencionales contra el cáncer, como la quimioterapia y la radioterapia, continúa produciendo beneficios terapéuticos significativos. Hasta el momento, se han obtenido resultados alentadores de extensos estudios clínicos preclínicos y humanos. Sin embargo, existe la necesidad de mejorar los vectores adenovirales para superar los problemas no resueltos que enfrenta este prometedor agente anticancerígeno, el principal de estos problemas es la respuesta inmune desencadenada por adenovirus y la falta de actividad en todas las células que conforman el tumor, lo cual amenaza su eficacia. Por lo tanto, mientras más avance el conocimiento de biología molecular, mayor será la eficiencia de la terapia y su correspondiente implementación [122][123].

Esta tesis tuvo como objetivo general atacar la heterogeneidad tumoral mediante la combinación de PETs, proponiendo PETs quimeras y promotores sintéticos para dirigir la actividad del primer gen del adenovirus, E1A. De esta manera, se busca reemplazar el promotor natural de E1A para que la actividad del adenovirus este dirigida transcripcionalmente y tenga actividad en la mayor cantidad de células posibles que afecten al tumor. Por un lado, los promotores híbridos buscan imitar el perfil de expresión génica de las distintas células tumorales a través de la incorporación de promotores de distinto origen que logren dirigir la terapia viral a un mayor número de células que conforman los subclones del tumor [65]. Por lo tanto, mientras más específicos sean los promotores a tejido y enfermedad, mayor será la eficiencia del tratamiento. Estos promotores proporcionan nuevas herramientas potentes y seguras para terapias génicas contra CCR [66]. Por otro lado, los promotores sintéticos son secuencias de ADN que no existen en la naturaleza, diseñados con el fin de regular actividad de genes de interés. Se busca a través

de la construcción del promotor sintético que los virus se repliquen específicamente en células cancerosas y por lo tanto la destrucción celular tumoral sea eficiente. Los diversos estudios sobre el armado de dichos promotores aporta información específica acerca del espaciamiento óptimo entre TF y TF, y con respecto a la caja TATA. A su vez, los diversos ensayos mostraron posibles combinaciones de factores conocidos, y la cantidad óptima de TF, mostrando que un número mayor a 6 TF no aumentaba la actividad del gen correspondiente [71]. Las secuencias reguladoras cis que conforman al promotor se integran como bloques de construcción, y pueden diseñarse mediante una estructura racional o ligadura aleatoria. El ensamblaje aleatorio de elementos ha demostrado un éxito prometedor [72]. Para ambos casos, mientras más conocimiento se tenga en el perfil de activación de los TF en las distintas etapas de la enfermedad, mejor serán las propuestas en los promotores encargados de dirigir la terapia. La hipótesis principal que promueve este trabajo de promotores híbridos es que la construcción de un promotor quimera de hasta 4 PETs cuya combinación de genes sobre-expresados en CCR podría agrupar en un solo virus la especificidad terapéutica para tratar CCR en tumores heterogéneos permite visualizar una opción a la terapia que permite tratar tanto un tumor primario como una metástasis en un mismo paciente. Con respecto a los promotores sintéticos, mientras más específicos sean los TFs incorporados en la secuencia, considerando los conocimientos previos sobre el armado de los mismos, mayor será la eficiencia de la terapia.

17. Análisis de datos

En este proyecto, se emplearon herramientas bioinformáticas a fin de obtener los TFs elementales en CCR, para poder así determinar la dirección transcripcional del adenovirus a través del armado de promotores híbridos y sintéticos. La determinación del perfil específico de TFs permite diferenciar a los promotores identificados en el laboratorio, para proponer de esta manera combinaciones posibles como promotores híbridos. En virtud de la diferencia observada en la expresión de las proteínas desarrollada en la sección de *Introducción*, se buscó establecer una metodología que permita establecer los factores asociados a dichas secuencias para explicar primero la sobre-expresión de dichas proteínas en CCR, y luego las similitudes y diferencias entre ellos. Para obtener los factores, se utilizaron las herramientas Match, Fimo y PROMO explicadas en la sección de *Materiales*, las cuales permitían ingresar una secuencia y como output presentaban los factores asociados a dicha secuencia ingresada. A su vez, cada factor estaba asociado con un valor de probabilidad, la matriz con la cual se reconoció el factor, la posición de inicio y final, y luego cada plataforma presentaba algunos resultados individuales. El hecho de poder ingresar una secuencia de interés era una característica fundamental en este trabajo ya que las secuencias de los promotores estaban previamente analizadas en el laboratorio y por lo tanto, la secuencia de interés no era toda la secuencia del promotor (las cuales suelen ser entre 2000 a 4000 pb). Luego, a través de la herramienta de RStudio, se formatearon las tablas para que la posición de los factores encontrados se posicionen en función del inicio de transcripción de cada promotor. Para ver las secuencias analizadas y los correspondientes inicios de transcripción, las mismas se presentan en la sección de *Anexo*. Una vez realizado este procesamiento, se normalizó la nomenclatura utilizada en las distintas plataformas para los TFs encontrados en función de HGNC. Habiendo realizado el primer filtro, eliminando así a los TFs que se encontraban después del inicio de transcripción, se implementó la herramienta ToppGene para el análisis de priorización génica, obteniendo de esta manera un enriquecimiento funcional

de las proteínas elegidas en el laboratorio, y los factores estadísticamente significativos para los distintos promotores. En el caso de VWA2, dado que no se obtuvo ningún factor estadísticamente significativo con esta metodología, se realizó un análisis directamente de los resultados obtenidos de las plataformas online originales. Las metodologías planteadas se reflejan en la figura 23 para promotores híbridos y 24 para promotores sintéticos. Una vez obtenidos los TFs, se realizó una clasificación en función de las variables definidas en 2 y 3, y la descripción de los mismos. Estos resultados se presentan en las tablas 13 y 14.

18. Factores de transcripción encontrados

Gracias a la metodología planteada, el primer resultado importante a destacar es que todos los promotores seleccionados por el laboratorio presentan TFs específicos de cáncer y de tejido que explican la sobre-expresión de las proteínas correspondientes a CCR con respecto al tejido normal. Si bien los perfiles de expresión de cada TF difiere, tal como se demostró con las figuras 25, 26, 27, 28, 29, 30, 31 y 32, todos los promotores presentan TFs con perfil de sobre-expresión en CCR y la diferencia hace a caracterización propia de cada promotor. A su vez, los factores encontrados coinciden con la bibliografía correspondiente. Diversos estudios han encontrado TFs asociados a A33, AQP5, CDC25A, CDC25B, GLI1, FSCN1, NQO1 y VWA2. En este trabajo no solo se corroboran dichos resultados, sino además se presentan nuevos TFs asociados a cada promotor. La tabla 15 presenta un resumen de dichos TFs para cada promotor. A continuación se analizan los ocho promotores seleccionados en función de los resultados obtenidos y la bibliografía.

Para **A33**, se encuentran los factores CDX2 y KLF4 en las primeras dos posiciones de la tabla. Esto se puede corroborar en la tabla 5. Teniendo en cuenta el perfil de expresión de CDX2, el cual es específico únicamente para CCR, la presencia de este factor explica la potencia del promotor en la enfermedad tanto en estadios primarios como más avanzados. Si bien se presentan solo 6 TF, esto ocurre ya que la secuencia de A33 es más corta que la de los otros promotores, y sobre todo de esos 6 TF, 5 de ellos están asociados a un rol específico en cáncer.

Para **AQP5**, lo primero a mencionar es que es uno de los pocos promotores que tiene un TF distinto a TP53 en la primera posición de genes rankeados por *ToppGene*. En las primeras tres posiciones se encuentran STAT3, NFKB1 y NR3C1, y en la cuarta se presenta TP53. Según un estudio realizado por Rachel Haviland et al., la sobreexpresión de STAT3 se encuentra asociado a la sobreexpresión de AQP5. Si bien la relación específica entre ambos genes sigue en discusión, Haviland detectan una regulación positiva entre ambos, lo cual explicaría la aparición de este TF en el promotor. A su vez, diversos estudios explican el bucle de retroalimentación positiva entre IL-6 / STAT3 / NF-kB que vincula la inflamación con el cáncer y mantiene las células en un estado transformado. Esto explica el ranking del primer y segundo gen en AQP5 [104][105]. Asimismo, tanto STAT3, como NF-kB1 y NR3C1 tienen un rol determinado en la proliferación celular, lo cual explica el rol de AQP5 de proliferación celular en el CCR. El factor RELA que aparece en la posición 15 forma un complejo heterodimérico RELA-NFkB1, el cual funciona como activador transcripcional (THPA). Por otro lado, factores encontrados tales como STAT1 y STAT5A tienen una actividad oncogénica a través de alteraciones transcripcionales o interacciones proteína-proteína, induciendo señalización antiapoptótica, respuesta e invasión aberrante del ADN, metástasis y la EMT. A su vez, estudios recientes han demostrado que la

señalización STAT6 es esencial para la EMT inducida por IL-4- e IL-13 y la agresividad de las células CCR [109]. Por otro lado, existe una regulación positiva entre FOXA1 y ESR1, lo cual explica la aparición de dichos factores en el promotor (THPA)[110]. En particular, ESR1 fue encontrado por las tres plataformas online en posiciones coincidentes, lo cual refuerza la validez de su posición. El mismo actúa sinérgicamente con NF-kB para activar la transcripción del gen que se encuentran regulando (tabla 14). A su vez, el análisis de muestras clínicas realizado por el grupo de Jiaoe Chen et al. 2018, de CCR no solo revela una correlación positiva entre E2F1, SP3 y STAT6, sino que el eje E2F1 / SP3 / STAT6 indica un mal pronóstico en pacientes con CCR, e incluso señala que la ruta E2F1 / SP3 / STAT6 es esencial para la EMT inducida por señalización de IL-4, explicando la agresividad de las células CCR [111]. Por otro lado, el factor NF1 encontrado en dos plataformas en la misma posición, se asocia con la desregularización de la red de señalización RAS en CCR encontrada en más del 70 % de los tumores, causada por una alteración en NF1[112]. Por lo tanto, la gran mayoría de los TFs hallados en el promotor AQP5 se encuentran asociados a EMT. Todos los TFs hallados con las respectivas posiciones se visualizan en la tabla 6.

Con respecto a **CDC25A**, ELK1 pertenece a la familia Ets de TFs y de la subfamilia del factor complejo ternario (TCF). Las proteínas de la subfamilia TCF forman un complejo ternario al unirse al factor de respuesta al suero y al elemento de respuesta al suero en el promotor del protooncogen c-fos. Por lo tanto, si bien no aparece el TF TCFB en la lista de ToppGene, si aparece el factor ELK1 encontrado por dos bases de datos en posiciones coincidentes, la cual a su vez coincide con el trabajo de (Antonio Iavarone et al; 1999). También se registran los factores c-FOS y c-JUN, lo cual coincide con la descripción del complejo ternario mencionada anteriormente [113]. Por otro lado, todos los TFs encontrados de la familia E2F (E2F1, E2F4, E2F5 y E2F7) fueron encontrados en la misma posición por distintas bases de datos, la cual coincide con el trabajo de (Antonio Iavarone et al; 1999): -62 a -55 [106]. A su vez, en la lista de TFs se encuentran SP1, AP-2, CEBPA, CEBPB y CEBPD. En todos los casos, las posiciones coinciden con el trabajo realizado por (Antonio Iavarone et al; 1999)[113].

Considerando el factor STAT3, el mismo se encontró en la posición “-317, -307” del promotor. STAT3, como TF, se identifica como un oncogén positivo que participa en el avance del tumor cuya activación inhibe la apoptosis a través del ciclo celular, la glucólisis aeróbica y la metástasis [107]. Además, STAT3 también afecta el microambiente tumoral mediante la estimulación de la angiogénesis y la inflamación y mediante la inmunosupresión mediadora [108]. Por lo tanto, la activación de STAT3 en las células tumorales puede promover la progresión tumoral. Recordando que este factor protéico actúa como regulador en la proteína CDC25A, el método empleado en este proyecto permitió no solo su detección en la secuencia, sino la posición del mismo en la secuencia.

MYB es un factor encontrado por dos plataformas en posiciones coincidentes, y una de esas posiciones coincide con la presentada por Antonio Iavarone (-65, -55 aproximadamente) [113]. c-Myb se sobreexpresa significativamente en los tejidos de CCR en comparación con los tejidos normales adyacentes y su alta expresión se correlaciona positivamente con metástasis de ganglios linfáticos y mal pronóstico. c-Myb promueve el crecimiento y la metástasis del CCR a través de la EMT inducida por c-fos [114].

Una de las claras limitaciones se refleja en los factores ausentes tales como c-myc. Esto se debe a ninguna de las tres plataformas online registró dichos factores, y por lo tanto, ToppGene no los tenía en la *lista de prueba*. Todos los TFs mencionados se visualizan en la tabla 7 junto con

las posiciones correspondientes.

Con respecto a **CDC25B**, TP53 se encuentra primero en la lista de factores, en la misma posición indicada por los investigadores en el trabajo de (M Dalvai et al; 2003). Por otro lado, tanto para el factor NFYB como SP3, la plataforma ToppGene los identificó con un valor-p ajustado mayor a 0.05 (0.5565 y 0.7082 respectivamente). A su vez, ETS1 se encuentra en las tres plataformas en posiciones coincidentes. Los factores con quienes interactúa descritos en la tabla 14 también se encuentran presentes. Los factores RARA y AP2 aparecen en dos plataformas en posiciones coincidentes. El factor TBP asociado a la caja TATA se encuentra por PROMO y FIMO en posiciones coincidentes. Esto se puede visualizar en la tabla 8.

Analizando los resultados de **FSCN1**, los dos factores principales hallados en la bibliografía se presentan en la tabla 9. En particular, CREB1 se encontró en las tres plataformas con posiciones coincidentes. A su vez, TP53 y NFKB1 son los factores más significativos hallados, ambos implicados en el progreso de la enfermedad. ATF2 también se encontró en dos plataformas. Algo importante a destacar es que FSCN1 presenta un TF (HOXB13) asociado únicamente a tejido intestinal. Según THPA, no se especifica un rol determinado en el CCR, y para Gepia, el mismo no tiene expresión en CCR, únicamente en cáncer de próstata. Sin embargo, diversos estudios argumentan un posible rol de HOXB13, teniendo un efecto supresor de tumores en el CCR, siendo un factor importante para la transformación de células colorrectales. Los altos niveles de expresión de HOXB13 se asocian con una supervivencia global prolongada en pacientes con CCR [125][126]. Estos estudios concuerdan con los resultados histológicos de FSCN1 presentados en la sección de *Introducción*, ya que FSCN1 se encuentra únicamente expresado en CCR, y no en tejido normal del colon y recto. Por lo tanto, la presencia de un factor asociado con el tejido y no con la enfermedad CCR podría indicar una expresión de la proteína en tejido normal, que no coincide con los resultados observados del perfil de expresión de dicha proteína. La descripción de todos los factores y su relación con CCR se encuentra en la tabla 14.

Con respecto a **GLI1**, FOXM1, el principal factor asociado a la regulación de GLI1 junto con NFKB1, fueron hallados por la plataforma PROMO. Ciertos factores tales como ETS1, CEBPA, SP1, SP3, MYB, AP-2, TFDP1 y IRF1 se presentan en dos plataformas, con posiciones coincidentes. Para una descripción de cada uno y su relación con CCR nuevamente se encuentra en 14. Sin embargo, los factores SOX9, CCND2 y CD133 no fueron encontrados. Los mismos no se encontraron por ninguna plataforma online, y por lo tanto, ToppGene no los presenta en el ranqueo genético ya que no están en la lista de prueba. Esto podría indicar, por un lado, que los TFs verdaderamente no se encuentran en el promotor GLI1, o por otro lado, que las plataformas elegidas no fueron lo suficientemente específicas para detectarlos.

Con respecto a **NQO1**, analizando los elementos necesarios para el complejo de TF ARE, se encontró NFE2L2 (Nrf2) y c-Jun, permitiendo la activación genética de NQO1. A su vez, los diversos estudios nombrados que establecen la relación con NFE2L2 como principal factor de regulación del promotor se corrobora con los resultados obtenidos, ya que este factor fue encontrado con un valor-p 4,17E-06. A su vez, el factor TP53 también se encontró, junto con los complejos AP-1 (FOS) y AP-2. A su vez, se encontraron diversos factores asociados a la caja TATA, y factores involucrados en la enfermedad de cáncer tales como PPARG, ESR2, XBP1, AR, PGR, IRF1, FOXA1, KLF4, NR1H2, BACH1, E2F1, STAT1, LEF1, HIC1, RARG, EWSR1, NF1, SP1, RARA, MAF, MYB, PAX5, ETS1, STAT6, GRB2, MECOM, MAZ, STAT4, YY1, e incluso con el factor HNF4A, el cual se ha demostrado que tiene un rol fundamental en la enfermedad CCR. Muchos de dichos factores actúan en sincronía con el complejo ARE y con

los factores asociados a XREs presentes en NQO1 según la bibliografía. La secuencia del núcleo ARE interactúa con los TFs bZIP (Jun, Fos, Fra, Nrf, Maf, Raf y NF-E2) [127]. Numerosos de estos factores fueron encontrados. Para más información, la tabla 11 presenta todos los TFs encontrados y 14 la descripción de los mismos.

Con respecto a **VWA2**, si bien el único TF encontrado hasta el momento es c-myc y el mismo no aparece en la tabla de factores para dicho promotor, c-myc se encontró aguas abajo del inicio de transcripción, en la posición 186. Por lo tanto, en el primer filtro aplicado quedó afuera de la selección final. Es importante aclarar que el factor si fue encontrado por las plataformas online, pero por la metodología de análisis implementada en este proyecto no fue tomado en cuenta. Por otro lado, si se detectan factores tales como TP53, POU2F1, FOXA1, MYB, FOXD3, FOS, SOX9, TBP (asociado a la caja TATA del promotor), HNF4A, entre otros, los cuales tienen roles determinados en la enfermedad. Estos factores son novedosos ya que no existen estudios actuales que busquen explicar la especificidad y potencia de la expresión de la proteína en CCR a través de los TFs de su promotor.

La tabla 14, la cual presenta la descripción de todos los TFs hallados permite hacer grandes análisis en cuanto a la explicación de cada factor en cada promotor. De esta manera, se explica muchas de las apariciones encontradas. A través de esta metodología empleada, se logró no solo reconocer la mayoría de los factores ya resaltados en los distintos trabajos, tal como se menciona en los párrafos anteriores, sino se lograron encontrar diversos factores asociados a la especificidad de tejido intestinal, a la enfermedad de cáncer o incluso a ambas características. Esa información se encuentra resumida en la tabla 15. Este proceso de investigación de los TFs de los ocho promotores y la especificidad y potencia que le confieren a cada uno de ellos permite predecir posibles puntos de corte para el armado de promotores híbridos. A su vez, haber identificado los roles de cada TF, teniendo en cuenta que vienen de proteínas diferencialmente expresadas de CCR, permite aportar información en cuanto a cuales de los ocho promotores seleccionar y a que secciones de los mismos para el armado de los promotores híbridos. Se realizó el análisis en cuanto a las similitudes y diferencias de los promotores seleccionados por el laboratorio desde una visión relacionada al perfil de TF de los mismos, y las conclusiones fueron relacionadas con el rol de la expresión de las proteínas correspondientes a los promotores. De esta manera, se desarrolló posibles combinaciones para el armado de promotores híbridos, con las secuencias correspondientes identificadas en el laboratorio y las secciones identificadas en las figuras 33, 34, 35, 36, 37, 38, 39 y 40.

Por otro lado, la identificación de los factores y su rol determinado en CCR, junto con la bibliografía encontrada, permitió un avance en la terapia adyuvante con una dirección transcripcional guiada por promotores sintéticos. Algunos de los posibles modelos se presentan en las figuras 43, 44 y 45. Actualmente, no hay trabajos científicos que reporten promotores híbridos o promotores sintéticos que direccionen transcripcionalmente a un AOL en CCR. Esta tesis busca ampliar el conocimiento existente en el área en una patología en particular. Si bien se proponen algunos modelos de promotores en este trabajo, el análisis es lo que se resalta. La presentación de los numerosos factores y sus perfiles de regulación permite proponer diversas propuestas al margen de las realizadas en este trabajo.

A modo de resumen, ambos promotores tienen el fin de direccionar transcripcionalmente al AOL-RC exclusivamente a las células malignas ya que su replicación depende de un promotor genético obtenido de genes sobreexpresados en forma selectiva en las células tumorales del CCR. Este trabajo se basó en la búsqueda de un AOL dirigido tanto por un promotor quimera como

con promotores sintéticos para lograr la eficacia terapéutica necesaria para el tratamiento de una patología con heterogeneidad intra e inter tumoral como lo es el CCR. Es importante mencionar que la cinética y la sensibilidad de la activación de los promotores propuestos se deben explorar aún más en los análisis *in vivo* que se realizan en el laboratorio. La bioinformática es una herramienta que nos permite conocer y estudiar posibles resultados para así afrontar mejor los resultados obtenidos en los análisis *in vivo*. Por lo tanto, estos resultados que provienen de una probabilidad estimada por plataformas online que implementan distintas bases de datos deben corroborarse con los análisis en el laboratorio para poder determinar la certeza de dichos resultados.

19. Perspectivas futuras

Teniendo en cuenta que la terapia génica es un campo que se encuentra en pleno desarrollo, esta tesis tiene como objetivo aportar información en el área de promotores para dirigirla transcripcionalmente al adenovirus oncolítico de manera más eficiente. Si bien en este trabajo se hace un estudio de los promotores seleccionados por el laboratorio (A33, AQP5, CDC25A, CDC25B, FSCN1, GLI1, NQO1 y VWA2) a partir de las secuencias de los mismos y así identificar los TFs asociados, las propuestas de promotores híbridos y sintéticos son solo el comienzo y aún se requiere de diversos estudios *in vitro* como estudios de PCR y PCR cuantitativo para verificar la presencia/ausencia de los TFs detectados en cada promotor, o por lo menos los principales asociados a las variables identificadas en el estudio y su correspondiente cantidad de ARN mensajero asociado a los transcritos. Se debe verificar la presencia de la caja TATA en todos los promotores, ya que la aparición de factores asociados a la misma no indica necesariamente que dicho promotor tiene una caja TATA asociada. A su vez, un ensayo de gen reportero para estudiar la regulación de la transcripción de las secuencias de interés. Un ensayo de doble híbrido permite probar dos proteínas a ver si interactúan entre sí. A largo plazo, los estudios *in vivo* van a permitir analizar respuestas inmunitarias para verificar el rol de los factores asociados a la actividad inmunitaria del hospedador.

En particular, en este trabajo se podrían haber realizado más ensayos en el área de bioestadística, como análisis de Kaplan-Meier y regresiones de Cox para verificar el análisis de sobrevida y los factores tanto de riesgo como de protección para el modelo. A su vez, se espera un análisis más detallado para identificar mejor el punto de corte de valores significativos. En este trabajo se consideró un p-valor de 0.05 como estadísticamente significativo, sin realizar pruebas de normalidad de las bases de datos. Se asumió que siendo plataformas establecidas hace tantos años, siendo reguladas y por lo tanto curadas, los datos numéricos de las mismas seguían una distribución normal. Un aspecto a mejorar es verificar esta distribución asumida.

Conclusiones

- El cáncer colorrectal (CCR) es la tercer neoplasia que causa mayor mortalidad según las cifras estimadas por la OMS en 2015. En Argentina, es el segundo cáncer con más incidencia luego del cáncer de mama, con 15.692 casos en 2018.
- Se desarrolló un proyecto para identificar promotores híbridos y sintéticos para desarrollar un tratamiento de la heterogeneidad de CCR con adenovirus oncolítico direccionado transcripcionalmente con los mismos.
- La búsqueda inicial de factores de transcripción (TFs) en las plataformas online Match, PROMO y Fimo permitió el análisis de los factores en las secuencias de los promotores de interés para el laboratorio. A continuación se listan la cantidad de factores distintos registrados en este primer análisis sin tener en cuenta las múltiples repeticiones de cada uno en diversas posiciones para A33, AQP5, CDC25A, CDC25B, FSCN1, GLI1, NQO1 y VWA2: 108, 214, 299, 158, 149, 364, 214, 158.
- Luego de un primer filtro aplicado, eliminando los factores que se encontraban en una posición aguas abajo de +1, los mismos se redujeron a 99, 196, 274, 145, 134, 229, 196, 94 respectivamente.
- El análisis de priorización genética con la herramienta ToppGene, empleando la *lista de entrenamiento* con los promotores seleccionados en el laboratorio, y la *lista de prueba* con los factores asociados a cada promotor, permitió reducir los factores de interés. Eliminando a aquellos que presentaban un valor-p ajustado mayor a 0.05, se redujeron a: A33 = 6, AQP5 = 61, CDC25A = 84, CDC25B = 63, FSCN1 = 72, GLI1 = 84 y NQO1 = 64.
- Para VWA2, que no se pudo implementar ToppGene como herramienta de filtro ya que no se encontró ningún factor estadísticamente significativo, se decidió trabajar con los valores de probabilidad asociados a cada plataforma. Dado que era de interés para el laboratorio, y en trabajos anteriores se ha demostrado la sobre-expresión de dicha proteína en CCR, se implementó una metodología que se ajuste a este promotor.
- Para promotores híbridos, con las variables definidas en este estudio, se emplearon herramientas de bioinformática de minería de texto sobre las bases de datos de *The Human Protein Atlas*, *Gepia*, *PubMed* principalmente. Los factores se redujeron a: 5, 32, 40, 33, 37, 45, 33 y 10 respectivamente.
- Para promotores sintéticos, con las variables definidas en este estudio, se emplearon herramientas de bioinformática de minería de texto sobre las bases de datos de *The Human Protein Atlas*, *Gepia*, *PubMed* principalmente. Para este campo no se asocia los factores

a los promotores, sino la importancia reside en los factores en sí y su correspondiente rol en la enfermedad.

- Se resalta la sobre-expresión de los factores CDX1, CDX2, ELF3, HNF1A, HNF4A, KLF5, NR1H2, POU2F2 y VDR en CCR.
- Se resalta la sobre-expresión del factor HOXB13 en el tejido intestinal.
- Se resalta la expresión de 96 factores asociados a la enfermedad del cáncer: AHR, AR, ARNT, ATF1, BACH1, CDX1, CDX2, CREB1, CTCFL, CUX1, DDIT3, E2F1, E2F4, E2F5, EGR1, EHF, ELF3, ELK1, EP300, ESR1, ESR2, ETS1, ETV1, ETV4, ETV5, EZH2, EWSR1, FLI1, FOS, FOXA1, FOXD3, FOXM1, FOXO1, FOXO3, GRB2, HIC1, HNF1A, HNF4A, HOXA13, IKZF1, INSM1, IRF1, ITGAM, c-JUN, KLF4, KLF5, KLF6, LEF1, LHX1, MAF, MAFB, MECOM, MEF2A, MNX1, MYB, MYC, MYOD1, NF1, NFATC2, NFE2L2, NFKB1, NR1H2, NR1H2, NR4A2, PAX5, PAX8, PBX1, PGR, PLAG1, POU2F2, PPARG, RARA, RARG, RB1, RELA, RREB1, SMAD1, SOX2, SOX5, SOX9, SP1, STAT1, STAT3, STAT4, STAT5A, STAT6, TAL1, TCF3, TEAD2, TFDP1, TP53, TP73, VDR, WT1, XBP1, YY1.
- Este análisis permitió esquematizar los promotores con sus correspondientes factores y así poder elegir distintas secciones de la secuencia para posibles combinaciones de los distintos promotores con las regiones de mayor interés de cada uno. Si bien en este trabajo se presentan algunas opciones de promotores híbridos y se sugieren las secuencias de trabajo de cada promotor, el campo se encuentra en pleno desarrollo y se pueden realizar distintas combinaciones para evaluar el correspondiente desempeño, distintas a las planteadas.
- Se plantean distintas combinaciones de promotores sintéticos a partir de la bibliografía y los resultados de los factores obtenidos. Nuevamente, siendo un campo en pleno desarrollo, las propuestas realizadas en este trabajo son a modo de ejemplo. Se pueden realizar distintas combinaciones y probar su correspondiente eficacia en la terapia.
- No se realizan análisis de bioestadística tales como ensayos de sobrevida que se podrían incorporar para enriquecer el análisis de los factores y los promotores elegidos en el laboratorio.
- La cinética y la sensibilidad de la activación de los promotores propuestos se deben explorar aún más en los análisis *in vitro* e *in vivo* realizados en el laboratorio. La bioinformática es una herramienta que nos permite conocer y estudiar posibles resultados para así afrontar mejor los resultados obtenidos en los análisis *in vivo*. Por lo tanto, este análisis que presenta resultados que provienen de una probabilidad estimada por plataformas online, las cuales implementan distintas bases de datos, deben corroborarse con los análisis en el laboratorio para poder determinar la certeza de dichos resultados.

Anexo

20. Secuencia de promotores

A continuación se presentan las secuencias de los ocho promotores seleccionadas por el *Instituto Leloir*. En todos los casos, se encuentra resaltado el inicio de transcripción del promotor, resultado de un análisis previamente realizado. En todas las figuras, el inicio de transcripción (+1) se encuentra marcado en azul.

```
|pagcaaatatgggcaacacccttcttttactagcttggaaacctacctctgagtgcatccctttttattattt|  
|atttctgtcagttataagagaggcctaccctttgtgagcagctctaggactttgtacacctgctaagtagggag|  
|aaggcaggggaggtggctggtttaaggggaacttgaggggaagttagggaagactcctctcgggacctttggagtag|  
|gtgacacatgagcccagcccagctcacctgccaatccagctgaggagctcacctgccaatccagctgaggctgg|  
|gcagaggtgggtgagaagagggaaaattgcagggacctccagttggggccaggccagaagctgctgtagctttaac|  
|cagacagctcagacctgtatggaggctgccagtg|
```

Figura 46: Secuencia del promotor A33 de 410 pares de bases. En azul: +1

```
|ccccgtcccgtcccgtcgacaaggggaaccccggcctgggagagggcgccctccgggatccggtgcctagtcca|  
|ggtactgcccagctaccgggctcgaggattgcgaagcgtcggggcaggctggcacggtgccacttttcccaaa|  
|acgccagccttccaagcccagaagctcgcggcggcccaggccgggagcggcccaagctcggacggcggaccgcct|  
|gcaggacccagcccggcggccggcccccggcggcgggtgagggaggtgagcggcggccagcctgcccggacgagc|  
|atcactccgacccagcccgggggtgaggggggtcaggatgctccggtcgcaggaggaaaaggaggagctggaccaa|  
|aagcccgaagagaagaaaaggggaagcggcgcacggagcgcggtaaaggccggcggagctagacgccccgaggt|  
|cggagtgaagcggcgggaccgagccccgtctcccagggagtcggggcgcaaggcaccgaggagagcggggagc|  
|caacctgggcgcatcatgcgagggcccgggacgctgggcccgtctacaccggcctgggtcacgtggcccggga|
```

Figura 47: Secuencia del promotor AQP5 de 600 pares de bases. En azul: +1

```
|ggacctccacaggttctccttagcctccacattgcctcagagtgtggggcgcccggctggggcgaggtagcgga|  
|ggcccaaagggggccgaagctaactggacggcagctcgcgatgggaactacgcttcccagcatgcgacggggcaa|  
|agggcctttcagccgcgagcagcgcctcgcaggttctgctgggagttttcattgacctctgctccccctctcat|  
|tttgatccccgctcttctgctctggctccgcccccttctgagagccgatgacctggcagagctcccgcgagccgc|  
|tttcttctccccctctcattggcccagcctagctgccattcggttgagaggaggagaagttgcttactgattggt|  
|ggattccgtttggcgccaactaggaagggggggcggggcagcagctggccccactgagccgctattaccgggaga|
```

Figura 48: Secuencia del promotor CDC25A de 450 pares de bases. En azul: +1

```
Egtgtctaacgcagaccgtacagcccagctgggttagcaaaacttcggggagccagttggagcctctccccatcc
ctagcgggtgatcccaggtgacgacatgccgggggggtcctgcggaggccaccctagggcggttgctgctgccttt
gggagtggtggagctccaaacatgtcgcgagagggcgattttgggaggccgggatcctcgcgccaggggatgtg
cgaggggtgtgggataaatcttaattcctccggcccacccaaagcctggaaatccagcctccgcgcctcttgccct
gccccccccgcctcagtcggccctcatctaaccgctaccccattggtggcgtccggggcgcggtgctgtt
atTTTTcgaatatataaggaggtggaagtggcagctgaaactagaggcttccctggctggtgctgagccccggcg
```

Figura 49: Secuencia del promotor CDC25B de 450 pares de bases. En azul: +1

```
CCAGGTCCTGGCCTCACCGGGTCCCCACCGCGCTGTGACCTCCAGCCTCTTCCCTCGTCTGCCTCTGGGCTTC
TGTAAGGCTATGTGCTCCAAGGCCACCTCCTCCAGGCAGCCCTCAGACCCCCACCTTCGCCAGTACCGATCTGC
ACCGTGGTCTCTGAAGTCTCCATCGTGACCTCAAACCTCGCTCGTCTTGTCTCCTAGCGAGGCTTGGGGTCGGGG
TGCCGAGGTGGGGGACATCCGGGGGGTTAGGTGGCTGGCGCGGGGAGCCGGGGTTGTGAGGGGTGATGTCCTC
AGGCGGCGCGCTGCGGGGTGCGGCGAGGACACCGGTGGGGTGAGAGCACCGGCGGGGCAGCAGCGGGGGCCGCA
GCGCCGGTCCCTCGGCCGGGGCCCTCCCGCGCGGAGCCAGGGGCGGGACAGGGGGCGTGGCCTGGTGGCGC
TGACGTCACCTCGCCTATAAAATGTCGGGGCGCGCTAGCTGGGCTTTGGAGCGCTGCGGAGGGTGCCTGCG
GGCCGCGGCAGCCGAACAAAGGAGCAGGGGCGCCGCGCAGGGACCCGCCACCCACCTCCCGGGGCGCGCAGCG
```

Figura 50: Secuencia del promotor FASCIN de 600 pares de bases. En azul: +1

```
GGCCTCTTGGGGCCAGCCGATCCAGAAATCCAAGCCGGGATTTAGTACTCACCAACAGCAGCGTGTTCAGCCGG
GGCGGGGGGGGGGGCGTAAGCAGTATAGGGTCCCTCAAGGGAGGGGGAGGATCCTGGGGGTCTGGGGGTGCAAT
AAGCCCGGCACCCCTTCTTGTCTCCAGCTACCCCGCCTCATCTCCAGAACGGCAAGAGGGAGGGAAATAGAA
GGGAGGTGAGGGGCGAGCGGGAAGAGCGGCGCGCCAGCGGCTGGAGAGAGAAAAAGTTTTGCAAAAGGGAA
AAAAAAGTTTGCCTTCTCGCGGGTGGTCCGGGCTTGCGGCCCGGCGGGCTGGGCCGGCGGGAGGGCTGGGGG
CAGGTTGGGGGGTGGGGTGGCATCGAGGCTGCGCTGCCGTGGCCCTCTCCGCCCCCTCCCCACCGCACACg
gcctctttggggccaACTCCAGCCCTGGACCGCGCATCCCGAGCCAGCGCCAGACAGAGGTGAGAAGGGGGGG
CAGGCGGGGACCACCTGGGAGCAGTGGGGAGGGGGCCTGAGGGGATGCTCAGCTTCTTAGGGACTCATCCAG
ACCC
```

Figura 51: Secuencia del promotor GLI1 de 604 pares de bases. En azul: +1

```

|gcgtgagccaccgcgctggtcccgttttggctattctttaatatgcttccctataactgctatctccacaccag
|ccttgccttagctcatgggaaacaaaaccaattagttccccattacctgacctgaggagcaggggtggtgcagt
|ggcatgcacccagggaaagtgtgttgtatgggcccccaagtgcagaatctgaatcttgcaaaggaagaaacaaatt
|cgtctccaaggagcatgtctcccaggactctcagccttccaaatccgcagtcacagtgactcagcagaatctga
|gcctagggcaccacagtaatccgcacccggctcttttctgctccatttctgtgccaaggctaaaagggcatgccc
|acttgatccctggactctcttgggacgacttccaccctgcatcctcttgcaacctcagggcacagtgccagatgg
|gcttgccttagcacccccagccagattttggagcctctgtcacacacaccctacaatcccctccccagcccc
|gagagacttttcttgacttccaccagttgctccggcggtgagagtggagaggccccctcctcatccccagget
|ccctcccttctggagctgcagcctcagcatcctccgcccagcaccaccagattcaggcgttgggtcccgcctt
|gtaggctgtccacctcaaacggggccgacaggatatataagagagaatgcaccgtgactacacagcgactccc
|acaaggttgagccggagccgcccagctcaccgagagcctagtccggccagggctgccccggcaaccacgagcc
|cagccaatcagcgcc
    
```

Figura 52: Secuencia del promotor NQO1 de 840 pares de bases. En azul: +1

```

|SCAATGAACAGGCATGCTCTGGCAGTTTTTAAATGAAGTATGGAAATTATGGATTGTGCAGTCATAGTCAACGTT
|CATTGAGAGCTTACTTTATATAAAGCATCATGCTCAGCATTGTGCTTTGTGGCATTGATTTTTTAAAAATCAA
|CGCCCTCCAATCTTAAGCAAGGATGATGGCCAAAAGGTCGGCTTATTACCCAAGTCAACAATAGGAGGGACAGG
|TAATTAAGGCAATTGCCATGACAAATAGAAAAGATGCCAAAGTCAACATTTTTTCAAGGAGCTATAAAGCCCG
|TGGAGGAGTCAGCTGACGGCACAGTgttcggcgtgcaccgctcggaggctgggtgaccgtagaagtgaagt
|ACTTTTTTATTTGCAGACCTGGGCCGATGCCGCTTTAAAAACGGAGGGGCTCTATGCACCTCCCTGGCGGTAG
|TTCTCCGACCTCAGCCGGTGGGTGGTCCCGCCCTCTCCAGGAGAGACAAACAGGTGTCCACGTGGCAGCC
|GCGCCCCGGGGCCCCCTCCTGTGATCCCGTAGCGCCCCCTGGCCCGAGCCGCGCCCGGGTCTGTGAGTAGAGCCG
|CCCGGGCACCGAGCGTGGTCCCGCTCTCTTCCG
    
```

Figura 53: Secuencia del promotor VWA2 de 632 pares de bases. En azul: +1

Bibliografía

- [1] Organización Mundial de la Salud. <https://www.who.int/es/news-room/fact-sheets/detail/cancer>. Consultado en: 1/04/2020
- [2] MARIAM JAMAL-HANJANI, SERGIO A. QUEZADA, JAMES LARKIN, AND CHARLES SWANTON, «Translational Implications of Tumor Heterogeneity», *Clin Cancer Res* 2015; 21(6) doi: 10.1158/1078-0432.CCR-14-1429.
- [3] XAVIER SAGAERT ARNO VANSTAPEL SANNE VERBEEK, «Tumor Heterogeneity in Colorectal Cancer: What Do We Know So Far?», *Pathobiology* 2018; DOI: 10.1159/000486721.
- [4] WAYTECK, L., BRECKPOT, K., DEMEESTER, J., DE SMEDT, S. C., AND RAEMDONCK, K., «A personalized view on cancer immunotherapy.», *Cancer Letters*, 2014; 352(1), 113–125.
- [5] SUZANNE L. TOPALIAN, M.D., F. STEPHEN HODI, M.D., JULIE R. BRAHMER, M.D., SCOTT N. GETTINGER. ET AL., «Safety, Activity, and Immune Correlates of Anti-PD-1 Antibody in Cancer», *N Engl J Med* 2012; 366:2443-2454, DOI: 10.1056/NEJMoa1200690
- [6] NATIONAL CANCER INSTITUTE: CANCER STAT FACTS: COLORECTAL CANCER. <http://seer.cancer.gov/statfacts/html/colorect.html> Consultado en: 12/05/2020
- [7] ESTADÍSTICAS - INCIDENCIA. <https://www.argentina.gob.ar/salud/instituto-nacional-del-cancer/estadisticas/incidencia> Consultado en: 13/05/2020
- [8] AMERICAN CANCER SOCIETY - ETAPAS DEL CÁNCER COLORRECTAL <https://www.cancer.org/es/cancer/cancer-de-colon-o-recto/deteccion-diagnostico-clasificacion-por-etapas/clasificacion-de-la-etapa.html> Consultado en: 15/05/2020
- [9] AMERICAN CANCER SOCIETY - CIRUGÍA PARA EL CÁNCER DE COLON <https://www.cancer.org/es/cancer/cancer-de-colon-o-recto/tratamiento/cirugia-del-colon.html> Consultado: 15/05/2020
- [10] AMERICAN CANCER SOCIETY -ABLACIÓN Y EMBOLIZACIÓN PARA EL CÁNCER COLORRECTAL <https://www.cancer.org/es/cancer/cancer-de-colon-o-recto/tratamiento/ablacion-emoblizacion.html> Consultado en: 15/05/2020

- [11] AMERICAN CANCER SOCIETY - QUIMIOTERAPIA PARA EL CÁNCER COLORRECTAL <https://www.cancer.org/es/cancer/cancer-de-colon-o-recto/tratamiento/quimioterapia.html> Consultado en: 15/05/2020
- [12] A T SOUGIANNIS, B N VANDERVEEN, R T ENOS, K T VELAZQUEZ, J E BADER, M CARSON, I CHATZISTAMOU, M WALLA, M M PENA, J L KUBINAK, M NAGARKATTI, J A CARSON, E A MURPHY, «Impact of 5 Fluorouracil Chemotherapy on Gut Inflammation, Functional Parameters, and Gut Microbiota», *Brain Behav Immun* 2019;80:44-55. doi: 10.1016/j.bbi.2019.02.020.
- [13] J. CALZAS RODRÍGUEZ, ET AL., «Capecitabina: un quimioterápico oral en la lucha contra el cáncer de mama y colorrectal metastásico», *Farmacía Hospitalaria* 2003; Vol. 27. Núm. 3. páginas 171-178
- [14] CHEMIOCARE - IRINOTECÁN, <http://chemocare.com/es/chemotherapy/drug-info/irinotecaacuten.aspx> Consultado en: 15/05/2020
- [15] EDUARDO VILAR AND STEPHEN B. GRUBER, «Microsatellite instability in colorectal cancer—the stable evidence», *Nature Reviews Clinical Oncology* 2010; volume 7, pages 153–162
- [16] CÁNCER COLORRECTAL: TIPOS DE TRATAMIENTO <https://www.cancer.net/es/tipos-de-c%C3%A1ncer/c%C3%A1ncer-colorrectal/tipos-de-tratamiento> Consultado en: 15/05/2020
- [17] LETICIA MOREIRA Y FRANCESC BALAGUER, «Inhibidores del EGFR: ¿qué pacientes se benefician de estos fármacos?», *Implicaciones clínicas de la investigación básica* 2010; Vol. 9 N.º 5
- [18] QUÉ ES LA INMUNOTERAPIA, Cancer.Net: Información al paciente aprobada por el médico American Society of Clinical Oncology (ASCO). <https://www.cancer.net/es/desplazarse-por-atencion-del-cancer/como-se-trata-el-cancer/inmunoterapia/que-es-la-inmunoterapia#:~:text=La%20inmunoterapia%2C%20tambi%C3%A9n%20denominada%20terapia,%20funci%C3%B3n%20del%20sistema%20inmunitario.> Consultado en: 15/05/2020
- [19] LEE S ROSEN, IRA A JACOBS, RONALD L BURKES, «Bevacizumab in Colorectal Cancer: Current Role in Treatment and the Potential of Biosimilars», *Target Oncol* 2017; 12(5):599-610. doi: 10.1007/s11523-017-0518-1.
- [20] DREW M. PARDOLL, «The blockade of immune checkpoints in cancer immunotherapy», *Nature Reviews Cancer* 2012; volume 12, pages 252–264.
- [21] BRUCE L LEVINE, JAMES MISKIN, KEITH WONNACOTT, CHRISTOPHER KEIR, «Global Manufacturing of CAR T Cell Therapy», *Mol Ther Methods Clin Dev* 2016 Dec 31;4:92-101. doi: 10.1016/j.omtm.2016.12.006.
- [22] AMERICAN CANCER SOCIETY - VACUNAS CONTRA EL CÁNCER Y SUS EFECTOS SECUNDARIOS <https://www.cancer.org/es/tratamiento/tratamientos-y-efectos-secundarios/tipos-de-tratamiento/inmunoterapia/vacunas-contr-el-cancer.html> Consultado en: 16/05/2020

- [23] ALAN MELCHER, KELLEY PARATO, CLIONA M ROONEY, JOHN C BELL, «Thunder and Lightning: Immunotherapy and Oncolytic Viruses Collide», *Mol Ther*, 2011; 19(6):1008-16. doi: 10.1038/mt.2011.65.
- [24] STEPHEN J. RUSSELL AND KAH-WHYE PENG, «Viruses as anticancer drugs», *Trends in pharmacological sciences*, 2007; 28(7): p. 326-333.
- [25] JOUNG-WOO CHOI, YOUNG SOOK LEE, CHAE-OK YUN, AND SUNG WAN KIM, «Polymeric oncolytic adenovirus for cancer gene therapy», *J Control Release*. 2015; 219: 181–191.
- [26] NASSER HASHEMI GORADEL, NASIR MOHAJEL, ZIBA VEISI MALEKSHAHI, SAMIRA JAHANGIRI, MASOUD NAJAFI, BAGHER FARHOOD, KEYWAN MORTEZAEI, BABAK NEGAHDARI, ARASH ARASHKIA, «Oncolytic adenovirus: A tool for cancer therapy in combination with other therapeutic approaches», *J Cell Physiol*. 2018;1-11, DOI: 10.1002/jcp.27850
- [27] BART THACI, ILYA V. ULASOV, DEREK A. WAINWRIGHT, AND MACIEJ S. LESNIAK, «The Challenge for Gene Therapy: Innate Immune Response to Adenoviruses», *Oncotarget*, 2011; Vol.2, No 3
- [28] D M NETTELBECK, V JÉRÔME, R MÜLLER, «Gene Therapy: Designer Promoters for Tumour Targeting», *Trends Genet* 2000; 16(4):174-81. doi: 10.1016/s0168-9525(99)01950-2.
- [29] JOUNG-WOO CHOI, JUNG-SUN LEE, SUNG WAN KIM, CHAE-OK YUN, «Evolution of oncolytic adenovirus for cancer treatment», *Advanced Drug Delivery Reviews*, 2012; 64(8), 720–729. doi:10.1016/j.addr.2011.12.011
- [30] ZEBIN MAO, SONG SHAN, YUNYAN ZHU, XIA YI, HUA ZHANG, YONGFENG SHANGN, AND TANJUN TONG, «Transcriptional regulation of A33 antigen expression by gut-enriched Kruppel-like factor», *Oncogene*, 2003; 22, 4434–4443.
- [31] CAMERON N. JOHNSTONE, SARA J. WHITE, NIALL C. TEBBUTT, FIONA J. CLAY, MATTHIAS ERNST, WILLIAM H. BIGGS, CARRIE S. VIARS, SUZANNE CZEKAY, KAREN C. ARDEN, AND JOAN K. HEATH, «Analysis of the Regulation of the A33 Antigen Gene Reveals Intestine-specific Mechanisms of Gene Expression», *The Journal of Biological Chemistry*, 2002; Vol. 277, No. 37, pp. 34531–34539.
- [32] J SAKAMOTO, H KOJIMA, J KATO, H HAMASHIMA, H SUZUKI, «Organ-specific Expression of the Intestinal Epithelium-Related Antigen A33, a Cell Surface Target for Antibody-Based Imaging and Treatment in Gastrointestinal Cancer», *Cancer Chemother Pharmacol*. 2000; 46 Suppl:S27-32. doi: 10.1007/pl00014045.
- [33] WEI WANG, QING LI, TAO YANG, GUANG BAI, DONGSHENG LI, QIANG LI AND HONGZHI SUN, «Expression of AQP5 and AQP8 in human colorectal carcinoma and their clinical significance», *World Journal of Surgical Oncology*, 2012; 10:242 <http://www.wjso.com/content/10/1/242>
- [34] TAO SHAN, XIJUAN CUI, WEI LI, WANRUN LIN AND YIMING LI, «AQP5: A novel biomarker that predicts poor clinical outcome in colorectal cancer.», *Biochemical and Biophysical Research Communications* 2014; 32: 1564-1570. DOI: 10.3892/or.2014.3377

- [35] LIZONG SHEN, ZHENGCAI ZHU, YIMING HUANG, YONGQIAN SHU, MAOCAI SUN, HAO XU, GUOQIANG ZHANG, RENHUA GUO, WEI WEI, WENXI WU, «Expression profile of multiple aquaporins in human gastric carcinoma and its clinical significance», *Biomedicine and Pharmacotherapy* 2010; 64 313–318. doi:10.1016/j.biopha.2009.12.003
- [36] PROFUMO V, GANDELLINI P, «MicroRNAs: Cobblestones on the Road to Cancer Metastasis», *Critical reviews in Oncogenesis* 2013; 18(4):341-55
- [37] SEYFRIED TN, HUYSENTRUYT LC, «On the origin of cancer metastasis.», *Crit Rev Oncog*, 2013; 18:43–73
- [38] FOLPRECHT GUNNAR, «Liver metastases in colorectal.», *Am Soc Clin Oncol Educ Book* 2016; 35:e186-92. doi: 10.1200/EDBK159185.
- [39] STEVEN R. ALBERTS , WILLIAM L. HORVATH , WILLIAM C. STERNFELD , RICHARD M. GOLDBERG , MICHELLE R. MAHONEY , SHAKER R. DAKHILRALPH LEVITT , KENDRITH ROWLAND , SURESH NAIR , DANIEL J. SARGENT , JOHN H. DONOHUE, «Oxaliplatin, Fluorouracil, and Leucovorin for Patients With Unresectable Liver-Only Metastases From Colorectal Cancer: A North Central Cancer Treatment Group Phase II Study», *Journal of Clinical Oncology* 2005; 23:9243–9249. <https://doi.org/10.1200/JCO.2005.07.740>
- [40] CARMINE STOLFI, DANIELE FINA, ROBERTA CARUSO, FLAVIO CAPRIOLI, MASSIMO CLAUDIO FANTINI, ANGELAMARIA RIZZO, MASSIMILIANO SARRA, FRANCESCO PALLONE AND GIOVANNI MONTELEONE, «Mesalazine negatively regulates CDC25A protein expression and promotes accumulation of colon cancer cells in S phase», *Carcinogenesis* 2008; vol.29 no.6 pp.1258–1266, doi:10.1093/carcin/bgn122
- [41] K GALAKTIONOV, AK LEE, J ECKSTEIN, G DRAETTA, J MECKLER, M LODA, D BEACH, «CDC25 phosphatases as potential human oncogenes», *Science*, 1995; Vol. 269, Issue 5230, pp. 1575-1577, DOI: 10.1126/science.7667636
- [42] SILVIA HERNÁNDEZ, LUIS HERNÁNDEZ, SÍLVIA BEÀ, MAITE CAZORLA, PEDRO L. FERNÁNDEZ, ALFONS NADAL, JAUME MUNTANÉ, CARME MALLOFRÉ, EMILIO MONTSERRAT, ANTONIO CARDESA AND ELÍAS CAMPO, «cdc25 Cell Cycle-activating Phosphatases and c-myc Expression in Human Non-Hodgkin's Lymphomas», *Cancer Research*, 1998; 58: 1762-1767.
- [43] WEIGUO WU, YOU-HONG FAN, BONNIE L. KEMP, GARRETT WALSH AND LI MAO, «Overexpression of cdc25A and cdc25B Is Frequent in Primary Non-Small Cell Lung Cancer but Is Not Associated with Overexpression of c-myc», *Cancer Research*, 1998; 58: 4082-4085.
- [44] DANIELA GASPAROTTO, ROBERTA MAESTRO, SARA PICCININ, TAMARA VUKOSAVLJEVIC, LUIGI BARZAN, SANDRO SULFARO AND MAURO BOIOCCHI, « Overexpression of CDC25A and CDC25B in Head and Neck Cancers», *Cancer Research*, 1997; 57: 2366-2368.
- [45] YANCHAO MA, RUOQIN WANG, HUIMIN LU, XIAOMI LI, GUANGBO ZHANG, FENGQING FU, LEI CAO, SHENGHUA ZHAN, ZHENXIN WANG, ZHONGBIN DENG, TONGGUO SHI,

- XUEGUANG ZHANG, WEICHANG CHEN, «B7-H3 promotes the cell cycle-mediated chemoresistance of colorectal cancer cells by regulating CDC25A», *Journal of Cancer* 2020; 11(8): 2158-2170. doi: 10.7150/jca.37255
- [46] ICHIRO TAKEMASA, HIROFUMI YAMAMOTO, MITSUGU SEKIMOTO, MASAYUKI OHUE, SHINGO NOURA, YASUHIRO MIYAKE, TAKASHI MATSUMOTO, TOMOHIKO AIHARA, NAOHIRO TOMITA, YASUHIRO TAMAKI, ISAO SAKITA, NOBUTERU KIKKAWA, NARIAKI MATSUURA, HITOSHI SHIOZAKI AND MORITO MONDEN, «Overexpression of CDC25B Phosphatase as a Novel Marker of Poor Prognosis of Human Colorectal Carcinoma», *Cancer Research*, 2000; Volume 60, Issue 11.
- [47] RIZENG LI, BINBIN WU, JIANFU XIA, LECHI YE, XIAOPING YANG, «Circular RNA hsa-circRNA102958 promotes tumorigenesis of colorectal cancer via miR-585/CDC25B axis», *Cancer Management and Research*, 2019; 9:11 6887–6893, DOI: 10.2147/CMAR.S212180
- [48] CHRISTINA KARLSSON, STEPHANIE KATICH, ANJA HAGTING, INGRID HOFFMANN, AND JONATHON PINES, «Landscape of transcriptome variations uncovering known and novel driver events in colorectal carcinoma», *The Journal of Cell Biology*, 1999; Volume 146, Number 3, 1999 573–583, <http://www.jcb.org>
- [49] GIOVANNA PIRA, PAOLO UVA ET AL., «Landscape of transcriptome variations uncovering known and novel driver events in colorectal carcinoma», *Nature Reports - Scientific Research*, 2020; 10:432, <https://doi.org/10.1038/s41598-019-57311-z>
- [50] YOSUKE HASHIMOTO, MAREK SKACEL, IAN C LAVERY, ABIR L MUKHERJEE, GRAHAM CASEY AND JOSEPHINE C ADAMS, «Prognostic significance of fascin expression in advanced colorectal cancer: an immunohistochemical study of colorectal adenomas and adenocarcinomas», *BMC Cancer* 2006; 6:241 doi:10.1186/1471-2407-6-241
- [51] YOSUKE HASHIMOTO, DAVID W. LOFTIS, JOSEPHINE C. ADAMS, «Fascin-1 Promoter Activity Is Regulated by CREB and the Aryl Hydrocarbon Receptor in Human Carcinoma Cells», *PLoS ONE*, 2009; 4(4): e5130. doi:10.1371/journal.pone.0005130
- [52] BARBARA M. PISKOR, ANNA PRYCYNICZ, EMILIA LUBOWICKA, KATARZYNA MINI-EWSKA, JUSTYNA ZINCZUK, KONRAD ZAREBA, KATARZYNA GUZINSKA-USTYMOWICZ2, «Immunohistochemical expression of Fascin-1 in colorectal cancer in relation to clinical and pathological parameters», *FOLIA HISTOCHEMICA ET CYTOBIOLOGICA* 2018; Vol. 56, No. 2, pp. 106–112. DOI: 10.5603/FHC.a2018.0011
- [53] AIDA U. JAWHARI, ANDREA BUDA, MICHELLE JENKINS, KHURRAM SHEHZAD, CATHERINE SARRAF, ‡ MASAO NODA, MICHAEL J. G. FARTHING, MASSIMO PIGNATELLI, AND JOSEPHINE C. ADAMS, «Fascin, an Actin-Bundling Protein, Modulates Colonic Epithelial Cell Invasiveness and Differentiation *in vitro*», *American Journal of Pathology* 2003; Vol. 162, No. 1.
- [54] CHENG ZHENG LIU, JIAN TAO YANG, JOON WON YOON, ELISABETH VILLAVICENCIO, KRISTINA PFENDLER, DAVID WALTERHOUSE, PHILIP IANNACCONE, «Characterization of the promoter region and genomic organization of GLI, a member of the Sonic hedgehog-Patched signaling pathway», *Gene* 1998; 1–11.

- [55] YIN-LU DING, YONG ZHOU, LEI XIANG, ZHI-PENG JI, ZHOU-HONG LUO, «Expression of Glioma-associated Oncogene Homolog 1 is Associated with Invasion and Postoperative Liver Metastasis in Colon Cancer », *International Journal of Medical Sciences.*, 2012; 9(5):334-338. doi: 10.7150/ijms.4553
- [56] YIN-LU DING, YONG ZHOU, LEI XIANG, ZHI-PENG JI, ZHOU-HONG LUO, «Gli and hedgehog in cancer: tumours, embryos and stem cells», *Nat Rev Cancer*, 2002; 361–372. <https://doi.org/10.1038/nrc796>
- [57] N. DAHMANE, J. LEE, P. ROBINS, P. HELLER AND A. RUIZ I ALTABA, «Activation of the transcription factor Gli1 and the Sonic hedgehog signalling pathway in skin tumours», *Nature* 1997; volume 389, pages 876–881.
- [58] BAHAR DEGIRMENCI, TOMAS VALENTA, SLAVICA DIMITRIEVA, GEORGE HAUSMANN AND KONRAD BASLE, «GLI1-expressing mesenchymal cells form the essential Wnt-secreting niche for colon stem cells.», *Nature* 2018; 558, 449–453.
- [59] LILI JI, YINGZE WEI, TAO JIANG, SHUYANG WANG, «Correlation of Nrf2, NQO1, MRP1, cmyc and p53 in colorectal cancer and their relationships to clinicopathologic features and survival», *Int J Clin Exp Pathol*, 2014; 4;7(3):1124-1131
- [60] ASMA CHINIGARZADEH, RAZAUDEN ZULKIFLI, IMAN YAZE AND REYHANEH RAHNAMAI TAJADOD, « Isolation And Cloning of Human NQO1 Promoter In Pgl3 Basic Vector», *INTERNATIONAL JOURNAL OF SCIENTIFIC and TECHNOLOGY RESEARCH*, 2012; VOLUME 1, ISSUE 7, ISSN 2277-8616.
- [61] KUOJUN ZHANG, DONG CHENG, KUN MA, XIAOXING WU, HAIPING HAO, AND SHENG JIANG, «NAD(P)H:Quinone Oxidoreductase 1 (NQO1) as a Therapeutic and Diagnostic Target in Cancer», *Journal of Medical Chemistry*, 2018; DOI: 10.1021/acs.jmedchem.8b00124
- [62] BAOZHONG XIN, PETRA PLATZER, STEPHEN P FINK, LISA REESE, ARMAN NOSRATI, JAMES KV WILLSON, KEITH WILSON AND SANFORD MARKOWITZ, «Colon cancer secreted protein-2 (CCSP-2), a novel candidate serological marker of colon neoplasia», *Oncogene* 2005; 24, 724–731. doi:10.1038/sj.onc.1208134
- [63] BEATRIZ GONZÁLEZ, FERRAN FECE DE LA CRUZ, JOHANNA KRISTINA SAMUELSSON, ANDREU ALIBÉS AND SERGIO ALONSO, «Epigenetic and transcriptional dysregulation of VWA2 associated with a MYC-driven oncogenic program in colorectal cancer», *Nature-Scientific Reports*, 2018; 8:11097, DOI:10.1038/s41598-018-29378-7
- [64] JAEIL KIM ET AL., «Molecular Imaging of Colorectal Tumors by Targeting Colon Cancer Secreted Protein-2 (CCSP-2)», *Neoplasia* 2017; 19(10):805-816. doi: 10.1016/j.neo.2017.07.003.
- [65] MIN GUAN, JUAN R RODRIGUEZ-MADOZ, PILAR ALZUGUREN, CELIA GOMAR, M GABRIELA KRAMER, STEFAN KOCHANEK, JESUS PRIETO, CRISTIAN SMERDOU, CHENG QIAN, «Development of optimized vectors for gene therapy», *Proc Natl Acad Sci U S A*. 1999; 96(2): 324–326. doi: 10.1073/pnas.96.2.324

- [66] KATARZYNA PIEKAROWICZ, ANNE T. BERTRAND, FÉRIEL AZIBANI, MAGDALENA MACHOWSKA, GISÈLE BONNE, RYSZARD RZEPECKI, «A Muscle Hybrid Promoter as a Novel Tool for Gene Therapy», *Proc Natl Acad Sci U S A.* 2019; DOI:<https://doi.org/10.1016/j.omtm.2019.09.001>, PlumX Metrics
- [67] L WU, J MATHERLY, A SMALLWOOD, J Y ADAMS, E BILICK, A BELLDEGRUN, M CAREY, «Chimeric PSA Enhancers Exhibit Augmented Activity in Prostate Cancer Gene Therapy Vectors», *Gene Ther.* 2001; 8(18):1416-26. doi: 10.1038/sj.gt.3301549.
- [68] O GRECO, B MARPLES, G U DACHS, K J WILLIAMS, A V PATTERSON, S D SCOTT, «Novel Chimeric Gene Promoters Responsive to Hypoxia and Ionizing Radiation», *Gene Ther.* 2002;9(20):1403-11. doi: 10.1038/sj.gt.3301823.
- [69] BH LIU, X WANG, YX MA AND S WANG, «CMV enhancer/human PDGF- β promoter for neuron-specific transgene expression», *Gene Ther* 2004; 11, 52–60 <https://doi.org/10.1038/sj.gt.3302126>
- [70] WHAT ARE SYNTHETIC PROMOTERS?, Synpromics – the leader in gene control and synthetic promoters. <https://www.synpromics.com/about-us/faq/what-are-synthetic-promoters> Consultado en: 3/02/2020
- [71] H. MOHAMED, Y. CHERNAJOVSKY AND D. GOULD, «Assembly PCR synthesis of optimally designed, compact, multi-responsive promoters suited to gene therapy application», *Scientific Reports* 2016; 6:29388 DOI: 10.1038/srep29388
- [72] MICHAEL L. ROBERTS, «The Use of Functional Genomics in Synthetic Promoter Design», *Computational Biology and Applied Bioinformatics* 2011; DOI: 10.5772/20653. <https://www.intechopen.com/books/computational-biology-and-applied-bioinformatics/the-use-of-functional-genomics-in-synthetic-promoter-design>
- [73] ILARIA MOGNO, FRANCESCO VALLANIA, ROBI D MITRA, BARAK A COHEN, «TATA Is a Modular Component of Synthetic Promoters», *Genome Res* 2010; 20(10):1391-7. doi: 10.1101/gr.106732.110.
- [74] G M EDELMAN, R MEECH, G C OWENS, F S JONES, «Synthetic Promoter Elements Obtained by Nucleotide Sequence Variation and Selection for Activity», *Proc Natl Acad Sci U S A.* 2000; 28;97(7):3038-43. doi: 10.1073/pnas.040569897.
- [75] JAESEOK HAN, BRIENNE MCLANE, EUNG-HWI KIM, JI-WON YOON, HEE-SOOK JUN, «Remission of Diabetes by Insulin Gene Therapy Using a Hepatocyte-Specific and Glucose-Responsive Synthetic Promoter», *Mol Ther.* 2011 Mar;19(3):470-8. doi: 10.1038/mt.2010.255.
- [76] TOLGA CAN, «Introduction to Bioinformatics», *Methods Mol Biol* 2014; 1107:51-71. doi: 10.1007/978-1-62703-748-8-4.
- [77] BAYAT, A, «Science, medicine, and the future: Bioinformatics.», *British Medical Journal*, 2002; 324(7344), 1018–1022. doi:10.1136/bmj.324.7344.1018
- [78] GREEN, E. D., WATSON, J. D., AND COLLINS, F. S., «Human Genome Project: Twenty-five years of big biology», *Nature*, 2015; 526(7571), 29–31. doi:10.1038/526029a

- [79] NATIONAL HUMAN GENOME RESEARCH PROGRAM - PANORAMA GENERAL DEL PROYECTO DEL GENOMA HUMANO <https://www.genome.gov/panorama-general-del-proyecto-del-genoma-humano> Consultado en: 1/06/2020
- [80] FARAHNAZ SADAT GOLESTAN HASHEMI ORCID ICON, MOHD RAZI ISMAIL, ET AL., «Intelligent mining of large-scale bio-data: Bioinformatics applications», *Journal - Biotechnology and Biotechnological Equipment*, 2018; 32:1, 10-29, DOI: 10.1080/13102818.2017.1364977
- [81] MARÍA CECILIA ARGIBAY, LUCIANA MONTIVERO, «Uso de herramientas bioinformáticas, bioestadísticas y modelos celulares en la búsqueda de potenciales biomarcadores del Cáncer de Endometrio», *Tesis*, 2019.
- [82] ALTMAN, R., «Current Progress in Bioinformatics 2016», *Briefings in Bioinformatics*, 2015; 17(1), 1–1. doi:10.1093/bib/bbv105
- [83] GRACIELA GONZALEZ, KEVIN BRETONNEL COHEN, CASEY S. GREENE, UDO HAHN, MARICEL G. KANN, ROBERT LEAMAN, NIGAM SHAH, AND JIEPING YE, «TEXT AND DATA MINING FOR BIOMEDICAL DISCOVERY», *Pac Symp Biocomput.* 2013; 368–372.
- [84] *Jaspar* <http://jaspar.genereg.net/docs/> Consultado en: 8/01/2020
- [85] XIAO LV, DANBO WANG, YUE MA, ZAIQIU LONG, «Analysis of the Oncogene BRAF Mutation and the Correlation of the Expression of Wild-Type BRAF and CREB1 in Endometriosis», *Int J Mol Med*, 2018; 41 (3), 1349-1356
- [86] DAVID ALFEGO, ULRICH RODECK, ANDRES KRIETE, «Global Mapping of Transcription Factor Motifs in Human Aging», *PLoS One*, 2018; 13 (1), e0190457.
- [87] EMILY L CASANOVA, ZACHARY GERSTNER, JULIA L SHARP, MANUEL F CASANOVA, FRANK ALEX FELTUS, «Widespread Genotype-Phenotype Correlations in Intellectual Disability», *Front Psychiatry*, 2018; 9, 535 2018 Oct 29 eCollection
- [88] TOMPA, M., LI, N., BAILEY, T. ET AL., «Assessing computational tools for the discovery of transcription factor binding sites», *Nat Biotechnol*, 2005; 23, 137–144. <https://doi.org/10.1038/nbt1053>
- [89] CALEB KIPKURUI KIBET, PHILIP MACHANICK, «Transcription factor motif quality assessment requires systematic comparative analysis», *F1000Research*, 2016; 4(ISCB Comm J):1429 Last updated: 01 APR 2016.
- [90] CHEN J, BARDES EE, ARONOW BJ, JEGGA AG., «ToppGene Suite for gene list enrichment analysis and candidate gene prioritization.», *Nucleic Acids Res.* 2009; 37(Web Server issue):W305-11. doi: 10.1093/nar/gkp427.
- [91] A E KEL, E GÖSSLING, I REUTER, E CHEREMUSHKIN, O V KEL-MARGOULIS, E WINGENDER, «MATCH: A Tool for Searching Transcription Factor Binding Sites in DNA Sequences», *Nucleic Acids Res*, 2003 Jul 1;31(13):3576-9. doi: 10.1093/nar/gkg585.

- [92] XAVIER MESSEGUER, RUTH ESCUDERO, DOMÈNEC FARRÉ, OSCAR NÚÑEZ, JAVIER MARTÍNEZ, M MAR ALBÀ, «PROMO: Detection of Known Transcription Regulatory Elements Using Species-Tailored Searchess», *Bioinformatics*, 2002 Feb; 18(2):333-4. doi: 10.1093/bioinformatics/18.2.333.
- [93] TIMOTHY L BAILEY, MIKAEL BODEN, FABIAN A BUSKE, MARTIN FRITH, CHARLES E GRANT, LUCA CLEMENTI, JINGYUAN REN, WILFRED W LI, WILLIAM S NOBLE, «MEME SUITE: Tools for Motif Discovery and Searching», *Nucleic Acids Res*, 2009; 37 (Web Server issue), W202-8.
- [94] CHARLES E. GRANT, TIMOTHY L. BAILEY, AND WILLIAM STAFFORD NOBLE, «FIMO: Scanning for occurrences of a given motifs», *Bioinformatics*, 2011; **27 (7)**, 1017-1018.
- [95] BENJAMINI AND HOCHBERG, «Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing», *Journal of the Royal Statistical Society*, 1995; 57: 289-300, DOI: 10.2307/2346101.
- [96] PETER J. THUL AND CECILIA LINDSKOG, «The Human Protein Atlas - A spatial map of the human proteome.», *Protein Science.*, 2017; 27(1), 233–244. doi:10.1002/pro.3307
- [97] PONTÉN F, JIRSTRÖM K, UHLEN M., «The Human Protein Atlas—a tool for pathology.», *J Pathol.*, 2008; 216(4):387-93. doi: 10.1002/path.2440.
- [98] TANG, Z., LI, C., KANG, B., GAO, G., LI, C., ZHANG, Z. «Gepia: a web server for cancer and normal gene expression profiling and interactive analyses.», *Nucleic Acids Research*, 2017, 45(W1), W98–W102. doi:10.1093/nar/gkx247
- [99] RIFQI ZAHROH JANATUNAIM, AZZANIA FIBRIANI, «Construction and Cloning of Plastic-Degrading Recombinant Enzymes (MHETase).», *Recent Pat Biotechnol.*, 2020; doi: 10.2174/1872208314666200311104541.
- [100] XINYUAN FENG, ZHIQI ZHANG, XIAOXIA LI, YAN SONG, JIANBANG KANG, DONGHONG YIN, YATING GAO, NAN SHI, JINJU DUAN, «Mutations in gyrB Play an Important Role in Ciprofloxacin-Resistant *Pseudomonas aeruginosa*», *Infect Drug Resist.*, 2019; 12:261-272. doi: 10.2147/IDR.S182272.
- [101] HGNC. <https://www.genenames.org/about/> Consultado en: 4/2/2020.
- [102] Z. HE, W. DONG, J. HU, X. REN, «AQP5 promotes hepatocellular carcinoma metastasis via NF-kB-regulated epithelial-mesenchymal transition», *Biochemical and Biophysical Research Communications*, 2017; doi: 10.1016/j.bbrc.2017.06.046.
- [103] WEI WANG, QING LI, TAO YANG, DONGSHENG LI, FENG DING, HONGZHI SUN, GUANG BAI, «Anti-cancer effect of Aquaporin 5 silencing in colorectal cancer cells in association with inhibition of Wnt/b-catenin pathway», *Cytotechnology*, 2018; <https://doi.org/10.1007/s10616-017-0147-7>
- [104] SITI-FAUZIAH HENDRAYANI, BOTHAINA AL-HARBI, MYSOON M. AL-ANSARI, GABRIELA SILVA, AND ABDELILAH ABOUSSEKHRA, «The inflammatory/cancer-related IL-6/STAT3/NF-kB positive feedback loop includes AUF1 and maintains the active state of breast myofibroblasts», *Oncotarget*. 2016; 7(27): 41974–41985.

- [105] BRADEN C MCFARLAND, SUK W HONG, RAJANI RAJBHANDARI, GEORGE B TWITTY JR, G KENNETH GRAY, HAO YU, ETTY N BENVENISTE, SUSAN E NOZELL, «NF- κ B-induced IL-6 Ensures STAT3 Activation and Tumor Aggressiveness in Glioblastoma», *PLoS One* 2013; 8(11):e78728. doi: 10.1371/journal.pone.0078728.
- [106] XUNDI XU, HIROFUMI YAMAMOTO, MASATO SAKON, MASAYOSHI YASUI, CHEW YEE NGAN, HIROKI FUKUNAGA, TETSUSHI MORITA, MINORU OGAWA, HIROAKI NAGANO, SHOJI NAKAMORI, MITSUGU SEKIMOTO, NARIAKI MATSUURA, AND MORITO MONDEN, «Overexpression of CDC25A Phosphatase Is Associated with Hypergrowth Activity and Poor Prognosis of Human Hepatocellular Carcinomas», *Clinical Cancer Research*, 2003; Vol. 9, 1764–1772.
- [107] HENG YANG, TAKAHIRO YAMAZAKI, FEDERICO PIETROCOLA, HENG ZHOU, LAURENCE ZITVOGEL, YUTING MA AND GUIDO KROEMER, «STAT3 Inhibition Enhances the Therapeutic Efficacy of Immunogenic Chemotherapy by Stimulating Type 1 Interferon Production by Cancer Cells», *Cancer Res*, 2015; 75(18) 3812–22.2015; DOI: 10.1158/0008-5472.CAN-15-1122
- [108] HUA YU, DREW PARDOLL AND RICHARD JOVE, «STATs in cancer inflammation and immunity: a leading role for STAT3», *Nature Reviews Cancer* 2009; volume 9, pages798–809, <https://doi.org/10.1038/nrc2734>
- [109] HUI CAO, JING ZHANG, HONG LIU, LEDONG WAN, HONGHE ZHANG, QIONG HUANG, ENPING XU, AND MAODE LAI, «IL-13/STAT6 signaling plays a critical role in the epithelial-mesenchymal transition of colorectal cancer cells», *Oncotarget*. 2016; 7(38): 61183–61198.
- [110] SANJIB CHAUDHARY, B. MADHU KRISHNA, AND SANDIP K. MISHRA, «A novel FOXA1/ESR1 interacting pathway: A study of Oncomine breast cancer microarrays», *Oncol Lett*. 2017; 14(2): 1247–1264.
- [111] JIAOE CHEN, CHAOJU GONG, HUIQIN MAO, ZHAOYUN LI, ZEJUN FANG, QIANG CHEN, MIN LIN, XIANG JIANG, YANYAN HU, WEI WANG, XIAOMIN ZHANG, XIANJUN CHEN, HONGZHANG LI, «E2F1/SP3/STAT6 Axis Is Required for IL-4-induced Epithelial-Mesenchymal Transition of Colorectal Cancer Cells», *Int J Oncol* 2018 Aug;53(2):567-578. doi: 10.3892/ijo.2018.4429.
- [112] TERJE AHLQUIST, IRENE BOTTILLO, STINE A DANIELSEN, GUNN I MELING, TORLEIV O ROGNUM, GURO E LIND, BRUNO DALLAPICCOLA, AND RAGNHILD A LOTHE, «RAS Signaling in Colorectal Carcinomas through Alteration of RAS, RAF, NF1, and/or RASSF1A», *Neoplasia*. 2008; 10(7): 680–686. doi: 10.1593/neo.08312
- [113] ANTONIO IAVARONE AND JOAN MASSAGUE, « E2F and Histone Deacetylase Mediate Transforming Growth Factor β Repression of cdc25A during Keratinocyte Cell Cycle Arrest», *MOLECULAR AND CELLULAR BIOLOGY*, 1999; Vol. 19, No. 1, p. 916–922.
- [114] XIAO QU, XUEBING YAN, CHENG KONG, YIN ZHU, HAO LI, DENGDENG PAN, XIAOHUI ZHANG, YONGQIANG LIU, FANG YIN, AND HUANLONG QIN, «c-Myb promotes

- growth and metastasis of colorectal cancer through c-fos-induced epithelial-mesenchymal transition», *Cancer Sci.* 2019; 110(10): 3183–3196.
- [115] LUCA BUSINO, MASSIMO CHIESA, GIULIO F DRAETTA AND MADDALENA DONZELLI, «Cdc25A phosphatase: combinatorial phosphorylation, ubiquitylation and proteolysis», *Oncogene*, 2004; 23, 2050–2056.
- [116] M DALVAI, O MONDESERT, B BUGLER, S MANENTI, B DUCOMMUN, C DOZIER, «Doxorubicin Promotes Transcriptional Upregulation of Cdc25B in Cancer Cells by Releasing Sp1 From the Promoter», *Oncogene*, 2013; 32(42):5123-8. doi: 10.1038/onc.2012.524.
- [117] CHUAN ZHANG, YONG WANG, YIFEI FENG, YUE ZHANG, BING JI, SEN WANG, YE SUN, CHUNYAN ZHU, DONGSHENG ZHANG, YUEMING SUN, «Gli1 promotes colorectal cancer metastasis in a Foxm1-dependent manner by activating EMT and PI3K-AKT signaling», *Oncotarget.*, 2016; 7:86134-86147. <https://doi.org/10.18632/oncotarget.13348>
- [118] ANI K. JAISWAL, «Human NAD(P)H:Quinone Oxidoreductase (NQO1) Gene Structure and Induction by Dioxin», *Biochemistry*, 1991; 30, 10647-10653
- [119] AMINAT S. OGUN; NEENA V. JOY; MENOUGH VALENTINE., «Biochemistry, Heme Synthesis», *StatPearls Publishing*, 2020; Bookshelf ID: NBK537329, PMID: 30726014
- [120] KAIMING WU, ZHENXIAN ZHAO, KUANZHI LIU, JIAN ZHANG, GUANGHUA LI, LIANG WANG, «Long noncoding RNA lnc-sox5 modulates CRC tumorigenesis by unbalancing tumor microenvironment», *Cell Cycle* 2017; 16(13):1295-1301. doi: 10.1080/15384101.2017.1317416.
- [121] KAIMING WU, ZHENXIAN ZHAO, KUANZHI LIU, JIAN ZHANG, GUANGHUA LI, LIANG WANG, «Coactivators and corepressors as mediators of nuclear receptor function: an update», *Mol Cell Endocrinol* 143 (1-2): 1-7. doi:10.1016/S0303-7207(98)00145-2
- [122] BEATRIZ GONZÁLEZ, FERRAN FECE DE LA CRUZ, JOHANNA KRISTINA SAMUELSSON, ANDREU ALIBÉS AND SERGIO ALONSO, «Epigenetic and transcriptional dysregulation of VWA2 associated with a MYC-driven oncogenic program in colorectal cancer», *Nature-Scientific Reports*, 2018; 8:11097, DOI:10.1038/s41598-018-29378-7
- [123] BART EVERTS, HENK G VAN DER POEL, «Replication-selective Oncolytic Viruses in the Treatment of Cancer», *Cancer Gene Ther.* 2005; 12(2):141-61. doi: 10.1038/sj.cgt.7700771.
- [124] TRUYEN NGUYEN, PAUL NIOI, AND CECIL B. PICKETT, «Replication-selective Oncolytic Viruses in the Treatment of Cancer», *J Biol Chem.* 2009; 284(20): 13291–13295. doi: 10.1074/jbc.R900010200
- [125] BINBIN XIE, BINGJUN BAI, YUZI XU, YUNLONG LIU, YIMING LV, XING GAO, FEI WU, ZHIPENG FANG, YING LOU, HONGMING PAN AND WEIDONG HAN, «Tumor-suppressive function and mechanism of HOXB13 in right-sided colon cancer», *Signal Transduction and Targeted Therapy* 2019; volume 4, Article number: 51.
- [126] C JUNG, R-S KIM, H ZHANG, S-J LEE, H SHENG, P J LOEHRER, T A GARDNER, M-H JENG, AND C KAO, «HOXB13 is downregulated in colorectal cancer to

confer TCF4-mediated transactivation», *Br J Cancer*. 2005; 92(12): 2233–2239. doi: 10.1038/sj.bjc.6602631

- [127] TRUYEN NGUYEN, PAUL NIOI, AND CECIL B. PICKETT, «Replication-selective Oncolytic Viruses in the Treatment of Cancer», *J Biol Chem*. 2009; 284(20): 13291–13295. doi: 10.1074/jbc.R900010200