Ph.D. Thesis in Biomedicine and Biotechnology

Role of *EDPR1* and the *ZNF518B* factor in the development of colorectal cancer

Francisco Gimeno Valiente

Supervisors

Dra. Josefa Castillo Aliaga

Dr. Gerardo López Rodas

Dr. Luis Franco Vera

Tutor

Dra. Nuria Paricio Ortiz

Department of Biochemistry and Molecular Biology

Faculty of Biology

University of Valencia

Valencia, September 2019

The co-supervisors of the present doctoral thesis, Dr Josefa Catillo Aliaga, Dr. Gerardo López Rodas and Dr. Luis Franco Vera,

CERTIFY:

That, **Francisco Gimeno Valiente**, graduated in Biology from University of Valencia, has carried out the present Doctoral Thesis "**Role of *EDPR1* and the *ZNF518B* factor in the development of colorectal cancer**" and that, to their judgement, meets all the necessary requirements in order to qualify for the **PhD in Biomedicine and Biotechnology**, for which purpose it will be presented at the University of Valencia. The work has been done under their direction, authorizing its presentation before the Qualifying Court. And for this purpose, this certificate is extended.

Valencia, September 2019.

Thesis Co-supervisors:

Dra. Josefa Casillo Aliaga	**Dr. Gerardo López Rodas**	**Dr. Luis Franco Vera**
Faculty of Biology,	Faculty of Biology,	Faculty of Biology,
University of Valencia	University of Valencia	University of Valencia

Thesis Tutor:

Dra. Nuria Paricio Ortiz

Faculty of Biology,

University of Valencia

“¡Corred, insensatos!”

Gandalf el Gris

Agradecimientos

Sin duda alguna es el apartado de la tesis más emotivo, y no es para menos. Me dispongo a escribir estas líneas de agradecimiento y me vienen a la mente recuerdos y personas increíbles que han pasado por mi vida a lo largo de estos 4 años.

Para empezar este apartado, quiero agradecer el apoyo y la profesionalidad de mis directores de tesis Luis, Gerardo y Pepa. Siempre me acordaré de esa entrevista en el despacho de Luis, en la que apostasteis por mí para poder formar parte del laboratorio. Muchísimas gracias, porque fue justo a partir de ese momento cuando empecé a conocer a la gente que formaría parte de mi vida hasta la fecha. Para mí, Luís, representas la sabiduría, el compromiso y el amor por la ciencia, solo te pido que no te cortes nunca la coleta, como tú dices, pues la ciencia perdería un valor importantísimo sin ti.

Cuando entré al laboratorio conocí a la persona que sería la piedra angular de mi tesis, y una referente profesional que me ha marcado para el resto de mi vida. Pepa no has sido solamente una directora más, te has convertido en mi madre científica y no puedo estar más orgulloso de poder tener a una directora como tú. En ese sentido, y así me consta ¡Soy la envidia de los doctorandos! Te doy las gracias por apostar por mí, por enseñarme todo lo que me has enseñado, por todos los consejos y por todas las risas, por tu capacidad de generar equipo, por valorar a las personas como se merecen, por todo y más… No puedo estar más agradecido por alguien que ha sacrificado incluso aspectos de su vida personal para poder realizar un buen trabajo junto a mí. Espero que algún día, no muy lejano, tengas el reconocimiento que tanto te mereces.

Agradecer por supuesto a Andrés Cervantes, me otorgaste la posibilidad de participar en tus proyectos, mejorando mis habilidades y aprendiendo más de lo que nunca hubiese pensado. Gracias por la oportunidad y por las facilidades.

No me puedo olvidar de dos grandes profesionales y amigas como sois Valentina y Noelia. Gracias a vosotras y a la oportunidad que me brindasteis en participar en vuestros proyectos, he podido tener una formación enorme en estos 4 años y he podido crecer científicamente junto a vosotras. Me habéis aportado infinidad de cosas positivas, no sé qué me deparará el futuro, pero espero que sea bien cerca de vosotras. También quiero agradecer al resto del equipo médico del grupo, pues indirectamente habéis sido cruciales en las investigaciones, a Marisol, Tania, Desam, Carolina…

Cuando volví de Londres me encontré con un laboratorio muy cambiado, y para bien, pues entró gente nueva maravillosa. Quiero agradecer a Federica, Fernanda y Manuel por el gran equipo que hacéis y por apoyarme moralmente en esta fase de la escritura.

Una parte crucial del laboratorio es el equipo de técnicos, quiero agradecer la paciencia que habéis tenido conmigo y el trabajo realizado a Ana, Cristina, a Zahara el último fichaje estrella, y por supuesto a Fany la *gobernanta* del labo.

También quiero que agradecer al resto de compañeros de los laboratorios 1, 2 y 3, a Sara, Iris, Ana, Ray, Carlos, Roberto… Pero también a las personas con las que empecé, no me puedo olvidar del apoyo, consejos y buenos momentos con gente que ya no está en INCLIVA, pero que han sido muy importantes para mí, a Ángela, Marce, Bea y a mi gran apoyo londinense a Lucía.

Por último, darle las gracias a mis padres y a Maite, que sin duda habéis sufrido la peor parte, horas y horas dedicadas al laboratorio que no he podido estar con vosotros, pero habéis comprendido que esto es mi pasión y es por ello que siempre me habéis apoyado. Sin vosotros directamente no habría llegado hasta aquí.

¡GRACIAS A TODOS POR TODO!

INDEX

Figure index

Table index

Abbreviations

5-FU	5-Fluorouracil
AJCC	American joint committee on cancer
APC	Adenomatous polyposis coli
ATM	ATM serine/threonine kinase
BCAM	Basal cell adhesion molecule
BSA	Bovine serum albumin
CAF	Cancer-associated Fibroblast
CDKs	Cyclin-dependent kinases
cDNA	Complementary DNA
CIMP	CpG island methylator phenotype
CIN	Chromosomal instability
CK	Cytokeratin
CMS	Consensus molecular subtypes
CRC	Colorectal cancer
CREB5	cAMP responsive element binding protein 5
Ct	Threshold cycle
DAB	3,3’-Diaminobenzidine
DDAH1	Dimethylarginine dimethylaminohydrolase 1
DMSO	Dimethylsulfoxide
DNMT	DNA dimethyl transferase
ECM	Extracellular matrix
EMA	European Medicines Agency
EMT	Epithelial to mesenchymal transition
FBS	Fetal bovine serum
FDA	Food and drug administration of the United States

FFPE	Formalin fixed paraffin-embedded
HAT	Histone acetyltransferases
HCV	Hospital Clínico de Valencia
HDAC	Histone deacetylases
hnRNPs	Heterogeneous ribonucleoprotein.
HSPA1A	Heat shock 70kDa protein 1A
LncRNA	Long non-coding RNA
miRNA	Micro RNA
MSI	Microsatellite instability
MNase	Micrococcal nuclease
NRP1	Neuropilin 1
pCMV	Human cytomegalovirus
PTBP2	Polypyrimidine tract binding protein 2
qPCR	Quantitative polymerase chain reaction
RBL2	Retinoblastoma-like protein 2
RT	Retrotranscription
S100A14	S100 Calcium Binding Protein A14
SCAN	Somatic copy number alteration
SD	Standard deviation
siRNAs	Silencing interfering RNAs
SNP	Single nucleotide polymorphisms
snRNPs	Small nuclear ribonucleoproteins
SRPX	Sushi repeat-containing protein
STR	Short tandem repeat
TAC	Transcriptome analysis console
TAM	Tumour-associated macrophages
TBST	Tris-buffered saline and TWEEN 20 mix

TCGA	The Cancer Genome Atlas
THBS1	Thrombospondin 1
TOMM34	Translocase of outer mitochondrial membrane 34
TSS	Transcription start site
UCIM	Central Research Unit of the Faculty of Medicine of Valencia

Brief abstract

Colorectal cancer (CRC) is one of the most common malignancies worldwide. Most of CRC related death is due to metastasis; therefore, finding molecular markers for prognosis of invasiveness, constitutes an appealing challenge. This work reports a significant increase in the expression of the *ZNF518B* and *EPDR1* genes and their major splicing isoforms in tumour tissues in a commercial cDNA array. The canonical isoforms were also up-regulated in tumour tissues of a prospective cohort of 101 patients of *Hospital Clínico de Valencia* when compared with adjacent tissues. *EPDR1* encodes a protein related to ependymins, glycoproteins involved in intercellular contacts. *ZNF518B*, encodes a putative zinc-finger transcription factor. These genes are only poorly known. After a thorough statistical model, this work describes that *EPDR1* up-regulation correlates with the TNM staging parameters, especially T and M. Functional analyses by silencing or overexpressing these genes in CRC cell lines highlight their role in invasiveness, dissemination and EMT of the cells. Both, expression and functional analysis support the inclusion of *EPDR1* and *ZNF518B* in panels of genes used to improve molecular subtyping of CRC. In this line, overexpression of *EPDR1* was observed in CMS3 and CMS4 subtypes, while the gene was under-expressed in CMS1. Next, the regulation of both genes in CRC cell lines was studied. Nucleosomal organisation over the transcription start site and the differential presence of the epigenetic marks H3K9ac, H3K27ac, H3K4me3 and H3K9me3 correlate with *ZNF518B* expression. Inhibition of histone deacetylases increases the transcription of the gene. Therefore, *ZNF518B*, apart from being a candidate for invasiveness prognosis in CRC, may be a target for epigenetic drugs. The expression of *EPDR1* is regulated by DNA

methylation and by miRNA. Present work opens the way to further studies to fully exploit the therapeutic potentiality of these genes in CRC.

Resumen breve

La principal causa de muerte por cáncer colorrectal (CCR), uno de los más comunes, es la metástasis. Así, encontrar marcadores de invasividad del tumor constituye un desafío importante. En este trabajo se muestra, primero con un *array* comercial de cDNA de CRC, que la expresión de *ZNF518B* y *EPDR1* y de sus dos isoformas principales de ayuste, aumenta significativamente en los tumores. Después, con muestras pareadas de una cohorte prospectiva de 101 pacientes del Hospital Clínico de Valencia (HCV), se observó que la expresión de ambos genes está elevada en el tumor. *EPDR1* codifica una proteína relacionada con las ependiminas, implicadas en contactos intercelulares y *ZNF518B*, codifica un putativo factor transcripcional. Un exhaustivo análisis estadístico correlacionó la sobreexpresión de *EPDR1* con los parámetros de estatificación TNM, especialmente T y M, en la cohorte del HCV. El silenciamiento o sobreexpresión de los genes han permitido demostrar que ambos intervienen en la invasividad y diseminación de las células tumorales y están implicados en la transición epitelio-mesenquimal. Ambos genes pueden, pues, incluirse en los paneles utilizados para establecer los distintos subtipos moleculares (CMS) de CCR. En los pacientes clasificados como CMS3 y CMS4, *EPDR1* se encuentra sobreexpresado mientras que ocurre lo contrario en pacientes CMS1. También se ha estudiado la regulación de ambos genes en líneas celulares de CCR. *ZNF518B* se regula por la organización nucleosomal y por la presencia diferencial de marcas epigenéticas H3K9ac, H3K27ac, H3K4me3 y H3K9me3 alrededor del sitio de inicio de la transcripción. La inhibición de histona desacetilasas aumenta la transcripción del gen, que, aparte de ser un buen candidato para el pronóstico de invasividad, puede ser una diana para fármacos epigenéticos. *EPDR1* está regulado por metilación del DNA y por miRNA. El presente trabajo abre

nuevas posibilidades para profundizar en la potencialidad terapéutica de ambos genes.

Resumen extendido

Introducción:

El cáncer colorrectal (CCR) es una de las neoplasias malignas más comunes en todo el mundo. A pesar de los avances de la medicina de precisión, la supervivencia a largo plazo de los pacientes con enfermedad metastásica avanzada sigue siendo pobre y se espera que la carga de CCR aumente en un 60% en la próxima década. Sin embargo, esta enfermedad es un buen candidato para los programas de detección, pudiendo mejorar el tratamiento y la supervivencia de los pacientes. La mayor parte de las muertes relacionada con el CCR se debe a metástasis tumorales y, por lo tanto, el hallazgo de marcadores moleculares para la predicción o el pronóstico de invasividad constituye un desafío atractivo, pues podrían agregarse a la colección disponible de potenciales marcadores. Esto es particularmente atractivo en vista del creciente interés en la mejora de la estadificación clínica del CCR. Esta razón, junto con los riesgos de diseminación metastásica, hace especialmente interesante la identificación de posibles dianas accionables en CCR. Estudios previos en nuestro laboratorio determinaron que la activación de la vía EGFR, incluida la activación constitutiva por la mutación KRAS G13D en líneas celulares de CCR, indujo modificaciones epigenéticas en hnRNP, proteínas directamente involucradas en el ayuste alternativo. Además, a través de análisis de RNA-seq de líneas celulares de CCR, se encontraron cambios en la distribución de ayuste alternativo, en función de la presencia de la mutación KRAS G13D, en un conjunto de genes entre los que se incluyen *EPDR1* y *ZNF518B*.

El nivel de expresión de estos genes difiere entre las distintas líneas celulares de CCR y la investigación actual ha sido llevada a cabo para investigar las posibles funciones de estos genes.

EPDR1 codifica una proteína relacionada con las ependiminas, una familia de glucoproteínas cerebrales. Las ependiminas son proteínas transmembrana que juegan un papel en los contactos intercelulares entre las células neurales. Al principio se sugirió que su dominio extracelular pudiera mostrar propiedades antiadhesivas y mostrar una capacidad dependiente de calcio para interactuar con las fibrillas de colágeno. El primer trabajo sobre la presencia de *EPDR1* en mamíferos se publicó en 2001. Nimmrich et al., donde en una búsqueda de genes que se expresaban de forma diferencial en tejidos tumorales, observaron una expresión significativamente elevada de *EPDR1* en dos líneas celulares de CCR, estando ausente en las células de mucosa normal. La secuencia del ADNc de *EPDR1* mostró cierta similitud con los genes codificantes de ependiminas, y se descubrió que el gen *EPDR1*, primeramente nombrado como *UCC1* (de sus siglas en inglés *Upregulated in Colorectal Cancer*) estaba significativamente sobreexpresado en dos de los tres tejidos tumorales humanos analizados de este trabajo, en comparación con los pareados de mucosa normal. Poco después, Kirkland y colaboradores informaron de la presencia de un gen relacionado con ependiminas altamente expresado en células hematopoyéticas, en algunos tejidos no hematopoyéticos y en varios tejidos malignos y líneas celulares. El gen, llamado en dicho trabajo como *MERP1* (de sus siglas en inglés *Mammalian Ependymin-Related Protein*), resultó ser el mismo que UCC1; ambos nombres ahora se usan como alias del gen *EPDR1*. Distintos trabajos publicados sobre la presencia de *EPDR1* en humanos se han ido sucediendo en la literatura a lo largo de estos últimos años. Dichos trabajos han descrito diferentes alteraciones en el nivel de

expresión o polimorfismos de un solo nucleótido en el locus *EPDR1* en una variedad de procesos patológicos o de desarrollo, que implican en la mayoría, si no en todos los casos, una disfunción de la adhesión celular.

Por otro lado, en relación al segundo gen a estudiar, los datos publicados sobre el papel y las propiedades de *ZNF518B* son muy limitados y no se ha demostrado ninguna participación directa en el cáncer hasta la fecha. Según los datos recuperados de Ensembl, el ayuste alternativo del gen *ZNF518B* da lugar a un total de cinco isoformas. Las isoformas 1 y 2 del gen son las más importantes y la abundancia de sus ARNm es aproximadamente similar en las líneas celulares que expresan el gen ZNF518B. La isoforma 1 y canónica codifica una proteína de 1074 residuos de longitud, mientras que la isoforma 2, por su parte, incluye un marco de lectura abierto que puede producir, hipotéticamente, una forma truncada del extremo N terminal de la proteína. Se debe destacar que no existen evidencias de la existencia real de esta proteína truncada en la actualidad. Las tres isoformas restantes no codifican para proteína. A partir de los datos a nivel de secuencia, se ha demostrado que la isoforma 1 codifica para una proteína de dedo de zinc, un supuesto factor de transcripción del cual, a nivel bibliográfico, no se habían descrito ni identificado sus posibles dianas. Aun así, existe literatura en la que se ha demostrado que ciertos polimorfismos de un solo nucleótidos producidos cerca o sobre el gen *ZNF518B* se han asociado con la gota y/o la concentración de urato en suero. Interesantemente para nuestro propósito, en un análisis proteómico del interactoma de G9A, Maier et al. descubrió que *ZNF518B* es uno de los genes que interactúan con dicha histona metiltransferasa. Es importante destacar que toda la proteína *ZNF518B* traducida in vitro, así como sus fragmentos de delecionados, pueden interactuar con G9A, lo que demuestra que múltiples dominios están involucrados en la interacción. Se ha

demostrado que *ZNF518B* regula positivamente la metilación de H3K9, lo que sugiere que su unión a G9A da como resultado la activación de la enzima. Estos hechos establecen una relación potencial entre *ZNF518B* y el cáncer, ya que diversos trabajos han demostrado la desregulación de G9A en varios tipos de cáncer. Concretamente, se ha observado que la enzima se sobreexpresa significativamente en los tejidos tumores de CCR en comparación con el epitelio normal pareado, además su regulación negativa inhibe la proliferación de células cancerosas. En vista de los datos anteriores, se podría sospechar que una sobreexpresión de *ZNF518B* podría tener efectos nocivos sobre la progresión del cáncer e invasividad. Sin embargo, hasta la fecha no existen trabajos sobre la expresión diferencial de *ZNF518B* entre tejidos cancerosos y normales.

Objetivos

En el marco de los antecedentes descritos anteriormente, planteamos resolver algunas preguntas sobre el posible papel de *EPDR1* y *ZNF518B* en CCR. Para resolver estas cuestiones, se propusieron tres objetivos principales para cada gen; primero, analizar el perfil de expresión de ambos genes en muestras de pacientes con CCR; segundo, estudiar el papel potencial de cada gen en CCR; tercero, investigar los mecanismos moleculares implicados en la regulación de la expresión de *EPDR1* y *ZNF518B* en CCR.

Métodos

La metodología requerida para la determinación de la expresión génica tanto en las diferentes líneas celulares de CCR como en muestras derivadas de pacientes ha sido la RT-qPCR, por su parte, técnicas como Western-Blot, inmunocitoquímica o inmunohistoquímica han sido requeridas para la valoración a nivel proteico. La RT-

qPCR ha permitido determinar el nivel de expresión de ARNm tanto de los genes completos como de las distintas isoformas de ambos genes generados por ayuste alternativo. Esta determinación ha sido valorada en tres cohortes distintas, una primera cohorte consistente en un *array* comercial de ADNc de muestras de CCR y muestras sanas; una segunda cohorte proveniente del Hospital Clínico de Valencia consistente en 101 pacientes de CCR, disponiendo de muestras tanto en fresco como incluidas en parafina; y una tercera cohorte *in silico* de la base de datos TCGA utilizada, en el caso de *EPDR1*, como cohorte de validación. Para estudiar la función de EPDR1 y ZNF518B, los genes fueron silenciado mediante ARNsi y/o sobreexpresados mediante plásmidos específicos en distintas líneas de CCR. Concretamente, se silenciaron los genes en las líneas celulares de CCR que poseen una expresión génica de *EPDR1* y *ZNF518B* elevada como son las DLD1 y HCT116. Por otra parte, la sobreexpresión se realizó sobre líneas que, o no expresan, o expresan el gen significativamente poco, como son las SW48 y las RKO. Tras el silenciamiento o sobreexpresión se examinaron las consecuencias derivadas de ello. Estas consecuencias se observaron mediante distintos ensayos y experimentos entre los que se comprenden ensayos de proliferación, migración, invasión, adhesión a colágeno de tipo I y determinación del nivel de marcadores de transición epitelio-mesénquima. Por otra parte, para estudiar la regulación de la expresión de los genes, se observó en el caso de *EPDR1*, la metilación de una isla CpG, determinándose cuantitativamente los valores de metilación mediante la plataforma MassARRAY. En el caso de *ZNF518B*, debido a la falta de islas CpG de las que se pudiese sospechar de su estado de metilación, se determinó la estructura de la cromatina situado alrededor del sitio de inicio de la transcripción, dicho procedimiento se llevó a cabo mediante un ensayo de protección de nucleasa micrococcal. Así mismo, se estudiaron

a nivel de nucleosoma las marcas epigenéticas más relevantes mediante el análisis por Nuc-ChIP. Para estudiar los efectos derivados de la disminución de *ZNF518B* en líneas celulares de CCR, se realizó un estudio del perfil transcriptómico. Para ello se requirió de un ensayo de *array* ClariomS, el cual contiene más de 20,.000 genes bien anotados.

El análisis estadístico se realizó en función del número de variables a comparar. La distribución de los datos se analizó mediante las pruebas de normalidad de Shapiro-Wilk y Kolmogorov-Smirnov. La igualdad de varianzas se analizó mediante la prueba de Levene. Todos los análisis estadísticos se llevaron a cabo con el software RStudio (versión 1.0.136) y GraphPad Prism6. Por consenso, todos los p-valores inferiores a 0,05 se consideran significativos. Cuando se realizó la comparación entre dos variables, se utilizó la prueba t de Student. Cuando el análisis implicaba la comparación de más de dos variables, se utilizó una prueba de Kruskal-Wallis o ANOVA de acuerdo con la normalidad de los datos.

Resultados

En el presente trabajo se describe un análisis de la expresión de *EPDR1* en diferentes cohortes de pacientes, concretamente en tres, siendo la primera un *array* de DNAc; la segunda, una cohorte de pacientes del Hospital Clínico de Valencia correctamente balanceada; y la tercera y última, una cohorte de validación *in silico* de datos procedentes de la base de datos TCGA. Destacar que la expresión del gen *EPDR1*, tanto el gen “total”, observado mediante sondas diseñadas en exones comunes en todas las isoformas, como en sus dos isoformas de ayuste alternativo principales, aumenta significativamente en tumores respecto a muestras normales de la primera cohorte. El *array* de ADNc comercial contiene 43 muestras de tumor de diferentes

estadios de la enfermedad, es decir, estadios I, II, III y IV o metastásicos y 5 muestras de mucosa normal sana. A pesar de observarse una tendencia en el aumento de la expresión de *EPDR1* en estadios IV, no se observa significatividad en el aumento de la expresión en los estadios metastásicos respecto a los estadios localizados o I, II y III. Se debe destacar que, en la segunda cohorte, tanto las isoformas canónicas, como la expresión total del gen, también se vieron aumentadas significativamente en las muestras tumorales respecto a las muestras pareadas sanas. Esta segunda cohorte, la cual consiste en una cohorte prospectiva de 101 pacientes con CCR del Hospital Clínico de Valencia (HCV), contiene determinado para cada uno de los pacientes tanto el estadio como la subclasificación en los parámetros T, N y M. Nuestros resultados experimentales se verificaron, con la tercera cohorte proveniente de bases de datos *in silico* disponible públicamente. Un modelo estadístico exhaustivo nos permite concluir que la regulación positiva se correlaciona con los parámetros de estadificación TNM, especialmente T y M, es decir, con la capacidad de invasión de las distintas paredes del colon, como es el parámetro T, y con la capacidad de formación de metástasis, como es el parámetro M. Por otro lado, la dependencia del tiempo de supervivencia de la expresión de *EPDR1* en tumores es cercana a la significancia, pero se observó una dependencia muy significativa con respecto a la expresión del gen en los bordes tumorales histológicamente normales. Se debe destacar, la valoración de cada una de las isoformas del gen *EPDR1* en muestras conservadas en fresco derivadas de pacientes con enfermedad localizada. En este caso, aparte de demostrar en muestras de pacientes que la expresión de la isoforma 1 es mayoritaria respecto a la 2, se pudo relacionar un aumento de la ratio de la isoforma 2/isoforma 1 con el parámetro T, siendo mayor la ratio en pacientes clasificados como T4, así mismo se ha podido demostrar una relación significativa

entre el estatus mutacional de KRAS con un aumento de la ratio entre las isoformas. Destacar la relación significativa entre el aumento de la ratio isoforma2/isoform1 con el aumento de la recaída en pacientes con CCR de la cohorte.

El papel potencial de la participación de *EPDR1* en esas características se estudió *in vitro* utilizando varias líneas celulares de CCR humano. Se usaron dos estrategias, la eliminación de *EPDR1* con ARNsi en células que expresan el gen, DLD1 y HCT116, y la sobreexpresión en líneas celulares que no expresan *EPDR1,* como son SW48 y RKO. De esta manera, se pudo observar que *EPDR1* aumenta la proliferación celular y promueve la migración celular, además se ha demostrado la interacción con las fibrillas de colágeno tipo I y la invasividad. Los resultados de ambos tipos de experimentos y ensayos, señalaron a nivel funcional un papel maligno de *EPDR1* en CCR. Por otra parte, se ha podido demostrar una correlación altamente interesante de la expresión de *EPDR1* con los diferentes subtipos moleculares. En este sentido, se ha podido observar que pacientes clasificados dentro de los subtipos moleculares CMS3 y CMS4 poseen niveles de expresión de *EPDR1* significativamente elevados. Por otra parte, aquellos pacientes clasificados dentro del subtipo molecular CMS1, los niveles de expresión de *EPDR1* se encuentran significativamente disminuidos. Debe destacarse, que los pacientes clasificados dentro del subtipo CMS3 se caracterizan por poseer un estatus mutacional de KRAS positivo, así como también aquellos con mutaciones en APC, curiosamente los pacientes clasificados dentro de CMS3 se caracterizan por poseer fenotipos hipermetilados. Los pacientes CMS4, por su parte, se caracterizan por poseer activadas las vías relacionadas con la transición epitelio-mesénquima. Por último, los pacientes clasificados dentro del CMS1, característicos por poseer mutaciones en

BRAF, se presentan como pacientes con fenotipos hipometilados. Estos resultados refuerzan el papel de EPDR1 observados *in vitro*.

Los estudios relacionados con la regulación de la expresión han podido revelar que la metilación del ADN en una isla CpG del primer exón traducible está implicada en la regulación de la expresión del gen *EPDR1*. Pudiéndose observar un aumento de la metilación significativo en líneas celulares que no poseen expresión, o se encuentra muy disminuida, del gen. Por su parte, líneas celulares como DLD1 y D-MUT1, línea celular isogénica de DLD1, poseen niveles de metilación en la isla CpG descrita significativamente bajos, siendo curiosamente más bajos en la línea celular D-MUT1. Estos experimentos fueron corroborados mediante un estudio de tratamiento con 5-azacytidina, en los que se observan una relación dosis-dependiente de la expresión de *EPDR1* en SW48 y RKO. Además, la regulación postranscripcional por microARN 193a-5p puede estar involucrada. Se ha observado un aumento del microARN 193a-5p en líneas celulares que no expresan el gen, en cambio la expresión del microARN 193a-5p es significativamente baja en HCT116 y DLD1. Curiosamente, la expresión del microARN 193a-5p es más baja en D-MUT1 que en un DLD1.

En base a la posibilidad de determinar experimentalmente el perfil mutacional de las muestras de CCR de la cohorte del HCV, ha sido posible relacionar el nivel de expresión de *EDPR1* con *KRAS* y otras mutaciones relacionadas con CCR detectadas en pacientes de dicha cohorte. Los pacientes se dividieron en dos grupos, de acuerdo con los niveles de expresión de *EPDR1*. Los pacientes con un nivel de expresión superior al promedio se clasificaron como "grupo de alta expresión", mientras que aquellos con expresión del gen por debajo del promedio se denominaron "grupo de baja expresión". Los pacientes portadores de mutaciones en *KRAS* y *APC* se

encontraron principalmente en el grupo de alta expresión, mientras que en los pacientes clasificados en el grupo de baja expresión hubo un enriquecimiento de mutaciones *TP53* y *BRAF*. Esto sugiere que las rutas KRAS y / o WNT son relevantes para la expresión de *EPDR1*.

Respecto al segundo gen a estudiar en el trabajo, destacar que el nivel de expresión de *ZNF518B* fue analizó en la primera cohorte, la cual consta de 45 muestras de ADNc derivadas de pacientes con CCR y 5 muestras de mucosa normal. Se observó que hay una expresión significativamente mayor en muestras tumorales, independientemente del estadio, ya sea I, II, III o IV, que en la mucosa normal. Respecto a la cohorte del HCV, formada por 101 pacientes, se pudo determinar la expresión de las isoformas mayoritarias de *ZNF518B* en muestras en fresco de pacientes con enfermedad localizada. En este caso, la isoforma 2 se encuentra significativamente más expresada que la isoforma 1, no pudiéndose correlacionar el aumento de la ratio con ningún parámetro clínico-patológico. Por su parte, en relación al gen total, se ha podido observar un aumento significativo de la expresión de *ZNF518B* en pacientes con CCR del HCV respecto a tejido sano.

Por otra parte, cumpliendo con los objetivos del papel funcional del gen y siguiendo las mismas estrategias *in vitro* descritas anteriormente para el gen *EPDR1,* se ha realizado tanto un silenciamiento del gen *ZNF518B* en líneas que lo expresan, como una estrategia de sobreexpresión mediante plásmidos en líneas celulares con falta de expresión. Los resultados del silenciamiento y la sobreexpresión del gen mostraron que *ZNF518B* favorece la migración e invasividad de las células tumorales, aunque su papel en la proliferación ha sido demostrado que no es relevante. Como en el caso de *EPDR1*, *ZNF518B* favorece la interacción de las células con las fibras de colágeno tipo I. Estas funciones sugieren un papel importante en la transición

epitelio-mesénquima, que se determinó mediante la observación de los marcadores correspondientes. Los resultados relacionados con este proceso demostraron que un silenciamiento de *ZNF518B* generaba una disminución significativa de N-cadherina a nivel de proteína, y una disminución de SNAIL a nivel de proteína y mensajero. Por su parte, se observó mediante RT-qPCR un aumento de la expresión génica de *CDH1* con el silenciamiento de *ZNF518B*.

Mediante un análisis transcriptómico, llevado a cabo comparando el perfil de expresión de líneas celulares de CCR no silenciadas con aquellas en las que *ZNF518B* había sido silenciado, reveló que un total de 347 genes se alteraron significativamente de acuerdo con el estadístico FDR (False Discovery Rate) <0.05. Muchos de estos genes están involucrados en diferentes procesos biológicos y vías relacionadas con la carcinogénesis. De entro estos procesos biológicos se encuentran la transición epitelio-mesenquima, la adhesión focal, el ciclo celular, la modificación de histonas, la apoptosis y la angiogénesis. Así mismo, se ha demostrado su relación con diferentes vías como la vía EGFR, la vía PI3KCA-AKT, la vía de las MAPK, la vía de WNT o la vía de RAS. Por lo tanto, *ZNF518B*, ya sea directa o indirectamente, puede regular muchos pasos de la transformación neoplásica.

Por otro lado, también se estudió la regulación de la expresión del gen *ZNF518B*. Se descubrió que tiene lugar a nivel de cromatina, ya que diferentes estructuras de cromatina y diferentes modificaciones epigenéticas de histonas de la región promotora, se correlacionaron con los niveles de expresión de *ZNF518B* en líneas celulares de CCR. Específicamente, la presencia diferencial de las marcas epigenéticas H3K9ac, H3K27ac, H3K4me3 y H3K9me3 se correlaciona con la expresión génica.

Discusión

El subtipo molecular basado en la expresión génica está ganando interés en la estratificación del cáncer y recientemente se ha enfatizado que las firmas transcripcionales pueden permitir un refinamiento del subtipo de CCR y, en este sentido, encontrar nuevos genes desregulados en la enfermedad, ya sea en el tumor mismo o en su medio ambiente, puede ayudar a construir paneles genéticos útiles para el diagnóstico, pronóstico o selección de terapia dirigida.

En la presente tesis se describe que el gen *EPDR1* está regulado por aumento en CRC. Este es el primer trabajo realizado con una gran cohorte de pacientes y verificado con *arrays* de ADNc, así como con el análisis bioinformático de bases de datos disponibles públicamente. Es de destacar que la regulación positiva de *EPDR1* está asociada con los parámetros T y M y, menos significativamente, con N. Esto está claramente de acuerdo con los resultados obtenidos después de silenciar y sobreexpresar el gen en las líneas celulares CCR. La difusión de las células cancerosas puede verse facilitada por los efectos de *EPDR1* sobre la adhesión a las fibras de colágeno tipo I, que se han descrito como las "carreteras" para la migración de células tumorales. Vale la pena señalar que la supervivencia de los pacientes depende más de la expresión de *EPDR1* en la mucosa adyacente no tumoral que en el tumor mismo. En consecuencia, debe enfatizarse la importancia de estudiar biomarcadores en tejidos adyacentes para predecir el riesgo de recaída.

También ha sido posible relacionar la expresión de *EDPR1* con la estadificación de CMS. Los pacientes con mayor expresión de *EPDR1* se clasifican dentro de CMS3, el subtipo molecular característico de las mutaciones en KRAS. También hay un grupo importante de pacientes con alta expresión de EPDR1 dentro

del subtipo CMS4, característico de pacientes con alteraciones en genes implicados en la transición epitelio-mesénquima. Esto concuerda con el papel de *EDPR1* en la transición epitelio-mesénquima presentado a lo largo de esta tesis. Los resultados mostrados pueden ser útiles para considerar la inclusión de *EPDR1* en paneles de genes utilizados para mejorar el subtipo molecular de CCR, haciéndolo más preciso.

Por otro lado, los resultados descritos en esta tesis muestran por primera vez que el aumento de la expresión *ZNF518B* en CCR respecto a muestras sanas. En cuanto a las isoformas *ZNF518B*, esta tesis también constituye el primer trabajo sobre su abundancia relativa en pacientes. La isoforma 1, como factor transcripcional que contiene dedos de zinc, es seguramente la isoforma que regula, directa o indirectamente, el gran conjunto de genes que se muestran a lo largo de la tesis. Además, es capaz de interactuar con la histona metiltransferasa G9A, lo que resulta en la activación de la enzima. G9A está involucrado en la invasividad y metástasis en varios tipos de cáncer. La isoforma 2, desprovista de los dominios de dedos de zinc, podría competir con la isoforma 1. Si esto ocurriera, los cambios en la proporción de isoformas representarían una forma adicional de regulación de la actividad de la metiltransferasa. Queda por determinar si los efectos de *ZNF518B* en la invasividad están mediados por la activación de G9A. Se ha reportado que G9A reprime la expresión de *E-cadherina* y el silenciamiento de *ZNF518B* da como resultado un aumento significativo de la expresión de *CDH1*. Por lo tanto, se puede sugerir que los efectos de *ZNF518B* en la migración celular están relacionados con la activación de la metiltransferasa G9A. A través de ensayos transcriptómicos, hemos visto que *ZNF518B* es un gen que regula la actividad de 64 genes.

Teniendo en cuenta que la regulación de la expresión del gen eucariótico debe considerarse dentro del panorama de la cromatina, el valor de los experimentos

descritos aquí se consolida, ya que sugieren de una forma contundente que la expresión de *ZNF518B* está regulada por la estructura de la cromatina y los factores epigenéticos. Este gen podría agregarse a la lista de dianas epigenéticas para la terapia del cáncer.

Conclusión

Los resultados actuales pueden ser útiles para considerar la inclusión de *EPDR1* y *ZNF518B* en paneles de genes utilizados para mejorar el subtipo molecular de CCR. El papel de los dos genes en el aumento de la capacidad de migración, invasión y adhesión de las células tumorales, es decir, en su malignidad, apunta al uso del nivel de expresión de *EPDR1* y *ZNF518B* totales y de sus isoformas como marcadores pronóstico para mejorar el diagnóstico de pacientes con CCR. Se requerirá un conocimiento profundo de ambos genes para explotar completamente el potencial terapéutico de *EPDR1* y/o *ZNF518B* para tratar el CCR con especial énfasis en el impacto sobre la metástasis y la invasión de tejido

INTRODUCTION

Introduction

1.1 Colorectal cancer

Cancer cells are characterized by dysregulation in the cell proliferation pathways, alterations in the response to cellular stress, evade growth suppressors factors, induction of angiogenesis and vascularisation, activation of the mechanisms of invasiveness and metastasis, resistance to cell death and the capacity of escape from immunological attack[1] [2]. All the processes discussed above are governed by genes, which can be used as biomarkers of disease progression. Currently, the research for new biomarkers is very relevant, since the early detection of cancer increases the probability of survival significantly. In addition, the detection of new biomarkers, at present, can serve to subclassify tumors in the new molecular classifications, more accurate than the classic histological ones[3].

The large intestine is shaped anatomically by different sections, the caecum, the appendix, the colon, the rectum and the anal canal[4]. The abnormal cellular proliferation in the intestinal tissue, called polyp or tubular adenoma, may progress to a tumour in the same walls of the colon or rectum, favouring its dissemination through the different blood or lymphatic vessels, generating a metastatic process[5]. When the carcinogenesis is performed in colon and rectal sections of the gastrointestinal tract, it is called colorectal cancer (CRC)[6]. CRC is, in more than 90% of cases, originated in the epithelial cells of the colorectal mucosa[7]. In lower percentages this phenomenon is developed in squamous cells and adenosquamous and spindle cells[7].

1.2 Incidence and epidemiology of the colorectal cancer

Worldwide, the CRC is the third type of cancer with the highest prevalence, and the fourth type in the number of death caused[6]. According to its incidence, it ranks as the second type of cancer, representing 13.2% and 12.7% in men and women respectively[8]. In Europe, the CRC represents the second cause of death in men, behind lung cancer[9]. In women, the CRC is in third position as the main cause of mortality after breast and lung cancer[9].

The age and the origin of the patients are important factors to highlight the incidence of the disease[10]. The incidence varies from 2-8%, under 40 years of age in Europe and the United State to 38%, 21%, 17% in Egypt, Saudi Arabia and Philippines respectively[5]. Emphasizing the importance of the worldwide distribution of the CRC, there is an incidence higher than 30% in European countries, Canada, Australia, New Zealand, Uruguay, Japan or South Korea, being less than 6% in most of countries belonging to the African continent or Asia. About mortality by CRC, more than 14% is observed in Russia, Eastern European countries, Spain, Argentina and Uruguay[6 11] (Figure 1). A positive correlation has been demonstrated in the incidence of CRC with the human development index of the countries studied[12].

Another key factor in the incidence of the disease is the age. More than 90% of the CRC cases occurs in patients with 50 years of age or older[9]. Despite of this, the incidence in people younger than 50 years continues increasing, incrementing too the poor prognosis of the disease due to its late detection, and therefore the presence of a disease in advanced stages[13].

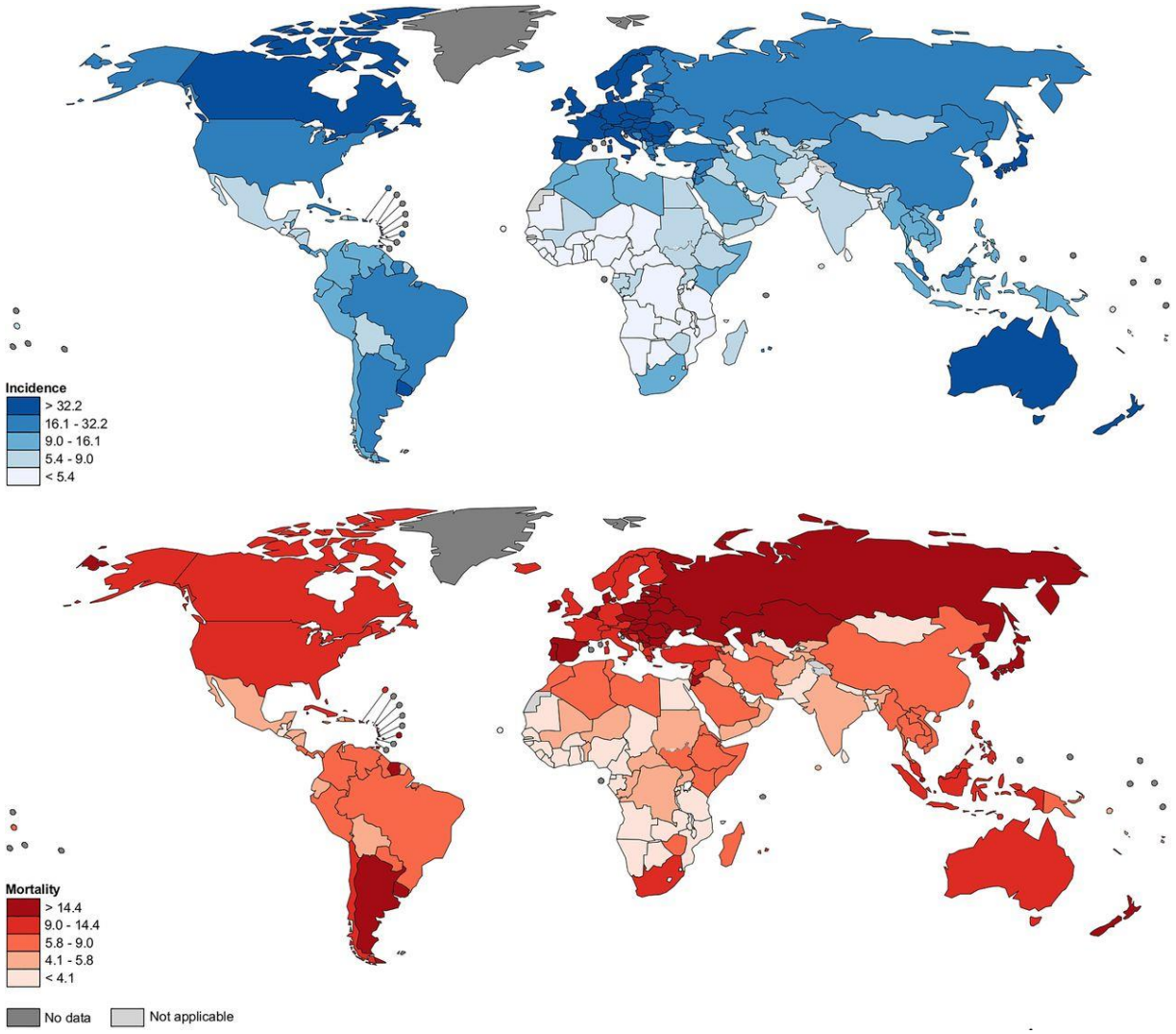

Figure 1: World map representative of the incidence (blue) and mortality (red) of the CRC by the World Health Organization (Arnold *et al*[6]).

1.3 TNM and molecular classification of the colorectal cancer

To know the tumour aggressiveness, a classification of the CRC has been developed based on the TNM characteristics (*T,* tumour size; *N,* lymph nodes affectation; *M,* appearance of distant metastasis) and the AJCC recommendations (American Joint Committee on Cancer)[14 15].

The AJCC classification orders the CRC evolution according to its degree of malignancy. In stage 0, cells with a high proliferative capacity are found on the layer closest to the lumen, mucosa or *muscularis mucosa*. In stage I, the tumour has penetrated to the next layers of the intestine, the submucosa or *muscularis propia*. In stage II, the tumour has invaded the outermost layers of the intestine, the subserosal or serous layer. The stage III is characterized by tumours that migrate through the different layers and there is a lymph node involvement affecting a certain number of them. The tumours classified as stages I, II or III are also known as localized tumours[16] [17]. Finally, in the stage IV or metastatic, the tumour has invaded any of the intestinal layers and an indeterminate number of regional lymph nodes, in addition to invasion of tissues or organs away from the colon or rectum, such as the liver, lung, ovary, or a distant lymph node to the place of the primary lesion[15] (Figure 2).

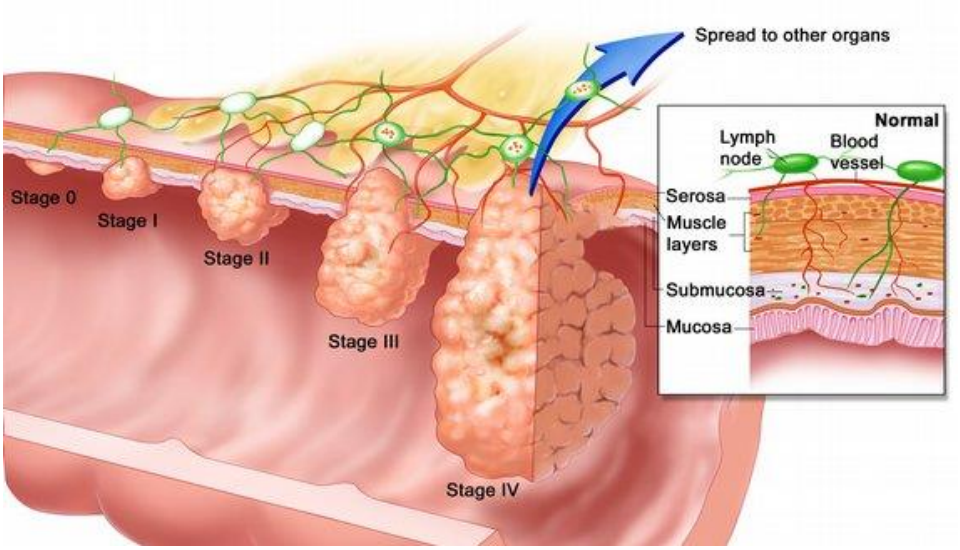

Figure 2: Development of CRC in relation with stages. In the inset the different layers that could be affected by the CRC can be seen. Image from the National Cancer Institute (NCI).

The TNM classification pays attention to some characteristics of the tumour[14] [18] [19]. The T parameter represents the degree of infiltration of the tumour by the different

intestinal layers, with T1 being the parameter with the least infiltration and T4 the highest. The N parameter represents the number of regional lymph nodes affected by the tumour, being N0 the absence of affected lymph nodes, N1 the infiltration to 1 to 3 regional lymph nodes, and N2 the infiltration to more than 4 lymph nodes[18]. Finally, the M parameter is indicative of metastasis outside the primary site of tumour growth, being M0 the absence of metastasis and M1 its presence[16 17].

Indeed, both classifications are highly related and widely used in clinical practice[20]. The AJCC stage I correspond to TNM stages named T1N0M0 and T2N0M0. In AJCC stage II there is a sub-classification, stage IIA corresponds to T3N0M0, while stage IIB to T4N0M0. The AJCC tumours classified within stage III, characterized by infiltration of tumours in regional lymph nodes, can also be sub-classified in IIIA, corresponding to T1N1M0 or T2N1M0, in IIIB, corresponding to T3N1M0 or T4N1M0, and IIIC related to any T, N2M0. Advanced or metastatic stages classified as AJCC stage IV is determined as any T, any N and M1[18] (Table 1).

Table 1: TNM Classification by AJCC stages of CRC

AJCC Stages		TNM Classification		
		T	N	M
Stage I		T1	N0	M0
		T2	N0	M0
Stage II	IIA	T3	N0	M0
	IIB	T4	N0	M0
Stage III	IIIA	T1	N1	M0
		T2	N1	M0
	IIIB	T3	N1	M0
		T4	N1	M0
	IIIC	Any T	N2	M0
Stage IV		Any T	Any N	M1

The AJCC classification of the CRC based on TNM parameters has several drawbacks since in many cases the patient's prediction accuracy is limited[21]. Therefore, other methods of tumour CRC cataloguing, based on the molecular classification, are increasingly being used for the staging of CRC tumours.

The molecular classification divides the tumour types into four molecular consensus subtypes (CMSs)[3]. The type CMS1, or immune microsatellite instability (MSI), represents 14% of cases, is characterized by an hypermethylated genome, with CpG islands methylation phenotype (CIMP)[22], *BRAF* mutations, activation of immune infiltration and worse survival after relapse[3] [23]. The CMS2 type or canonical[3], corresponds to 37% of tumours, is characterized by a high somatic copy number alterations (SCAN)[24] and activation of *WNT* and *MYC*[25] [23]. The CMS3 or metabolic[3], represents 13% of tumours, is characterized by a mixed MSI and low SCAN and CIMP, as well as *KRAS* mutations and a metabolic dysregulation[3] [25]. Finally, the CMS4 or mesenchymal, corresponds to 23% of tumours[25], have a high SCAN together with an activation of *TGF-β*, high stromal infiltration and angiogenesis. CMS4 subtype is related with the worse overall survival of the patients [3] [23] (Figure 3). This molecular and genetic classification could provide better clinical and biological information than the TNM classification, helping the physician to take clinical decisions.

CMS1 MSI immune	CMS2 Canonical	CMS3 Metabolic	CMS4 Mesenchymal
14%	37%	13%	23%
MSI, CIMP high, hypermutation	SCNA high	Mixed MSI status, SCNA low, CIMP low	SCNA high
BRAF mutations		*KRAS* mutations	
Immune infiltration and activation	WNT and MYC activation	Metabolic deregulation	Stromal infiltration, TGF-β activation, angiogenesis
Worse survival after relapse			Worse relapse-free and overall survival

Figure 3: Summary of the proposed molecular classification. Guinney *et al*[3].

1.4 Genetic alterations in the development of colorectal cancer

The genetic and epigenetic dysregulation of a canonical cell mechanisms can cause a potential cancer cell[1]. There are many ways in which these alterations can be generated.

In the colon, the evolution of normal epithelial cells to adenocarcinoma follows a predictable progression of histological and concurrent epigenetic and genetic changes[26]. As shown in the figure 4, there are two ways of developing CRC, the classical pathway (top) involves the development of tubular adenomas that can progress to adenocarcinoma. An alternate pathway that involves serrated polyps (at the bottom) appears to have the potential to transform into CRCs through a hyperplastic polyp. As can be seen in the figure 4, the normal epithelium takes its first steps towards the conversion of adenocarcinoma through the mutation of genes such as *APC* or *CTNNB1*, which are part of the so-called WNT pathway. The sequential acquisition of mutations shown in the figure 4 can take 10 to 15 years, but can progress much faster, being the case of patients with Lynch syndrome, who have germinal mutations in the DNA mismatch repair genes as *MLH1*, *MSH2*, *MSH6* or *PMS2*[27]. The second step of acquired mutations occurs in genes related to MAPK signalling, while the third one involves mutations in genes related to PI3KCA and TGFβ signalling. Finally, mutations related to the p53 signalling are acquired[26].

As commented above, several signalling pathways have been related to CRC growth and metastasis. One of the better-known pathways is that initiated by the activation of the epidermal growth factor receptor gene (*EGFR*). The overexpression of this receptor is observed in 60% of tumours and is associated with poor patient prognosis [29]. EGFR is located in the plasma membrane, containing an extracellular

domain, where EGF or TGFα, among others ligands binds, a transmembrane domain and an intracellular domain, where tyrosine kinase activity is found[30].

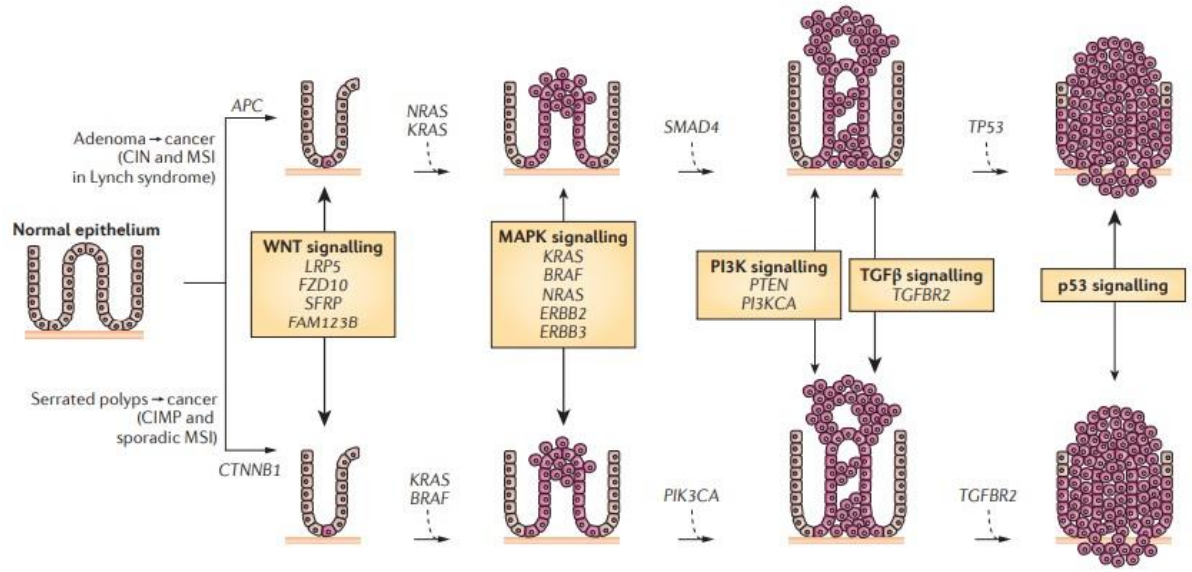

Figure 4: Diagram of the transformation from normal tissue to colorectal cancer. Both sequences comprise the progression of normal colon epithelial cells to aberrant crypt foci, followed by early and advanced polyps with subsequent progression to early cancer and then advanced cancer. The signalling pathways altered during the progression sequence are shown. (Kuipers *et al*[28])

EGFR, also known as ErbB1/HER1, is the prototype of the EGFR family that also includes HER2, HER3 and HER4. Based on its role in promoting cell proliferation and avoiding apoptosis, the EGFR has been defined as a proto-oncogene. In the case of the epidermal growth factor receptor 2 (HER2) its overexpression is usually determined by a genetic amplification. Recent studies have shown positivity in HER2 in approximately 2% of patients with CRC[31]. The HER2 protein has, as EGFR and the other two members of the epidermal growth factors family, HER3 and HER4, tyrosine kinase activity[32] [33]. Despite belonging to the same family note that HER2, unlike EGFR, is not activated by ligand binding, but by heterodimerization with another ligand binding receptor of the same family[34].

The overexpression of *EGFR* or other members of their family generates activation of different pro-oncogenic signalling transduction pathways, including RAS/RAF/MEK/ERK MAPK and PI3K/AKT/mTOR that promote the proliferation and survival of cells[35]. There are two types of strategies to inhibit this type of receptors in the clinical CRC treatment, as it will be better described in another section. In one hand, EGFR inhibition by monoclonal antibodies prevents binding of ligands in the extracellular domain. On the other hand, the inhibition of the tyrosine kinase domain means intracellular phosphorylation of the receptor[36].

Another important family of genes related to colorectal cancer, are RAS family genes. The *RAS* family is composed of the *HRAS, KRAS* and *NRAS* genes that encodes proteins with GTPases activity and are activated downstream of the epidermal growth factor receptors[37]. *KRAS* is known to be a highly mutated gene, found altered in 42% of patients with CRC, whereas *NRAS* is mutated in 9% of CRC patients[38]. Although these three genes are the most important, due to their involvement in cancer development, there are many other genes belonging to the *RAS* family, such as *MRAS, RRAS, RRAS2*.

The genes of the *RAS* family activate members of the following *RAF* family genes like *ARAF, BRAF* or *RAF1* genes. *BRAF* also have an important role in the development of CRC[39].

The activation of *RAF* family genes causes activation of downstream proteins, MEK and ERK. In turn, ERK activates transcription factors JUN, FOS, ELK, that modulate gene expression favouring cell processes like proliferation, angiogenesis, apoptosis, differentiation and metastases[38 40] (Figure 5).

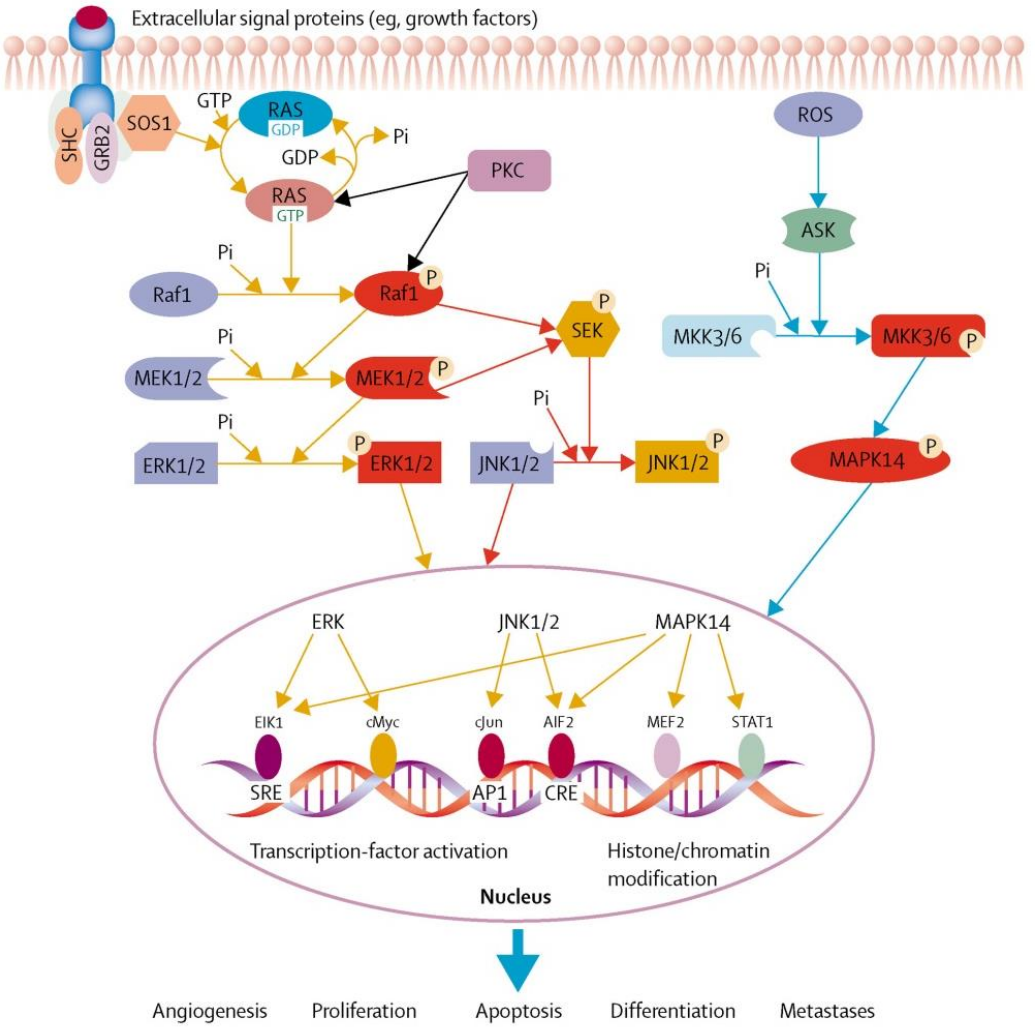

Figure 5*:* Serine-threonine kinase are involved in several signal transduction cascades activated by growth factors, stress or inflamation. Activation of RAS / RAF / MEK / ERK signalling pathway showing activation of different transcription factors and resulting in an increase in angiogenesis, proliferation, apoptosis, differentiation and metastasis (Fang JY and Richardson[40]).

Another of the most frequently dysregulated signalling pathways in cancer is PI3K/AKT/mTOR [41] [42] [43]. The overexpression of the PI3K/AKT/mTOR signalling has been reported in various forms of cancers, especially in colorectal cancers (CRC). PI3K is a member of lipid kinases family, which phosphorylates PIP2 and transforms it into its active state PIP3, favouring the activation of AKT [44]. *PTEN*, a well-characterized tumour suppressor gene, acts as a negative regulator of PI3K/AKT. Also

has a remarkable role in cell survival[45] and in the proliferation, inhibiting the transformation of PIP2 to the active form [44]. However, despite the importance of *PTEN* at this biochemical level, its importance as a prognostic factor in the clinic in patients with CRC has not been demonstrated [46].

Continuing with the PI3K/AKT pathway, the activation of AKT kinase favors the phosphorylation of TSC2 and its dissociation from TSC1[44]. This phosphorylation and the consequent TSC1/TSC2 dissociation causes *mTORC1* complex activating, and drives translation activation, lipid synthesis, nucleotide synthesis and decrease of autophagy[47]. On the other hand *mTORC2* complex favours the activation of *AKT*[42] (Figure 6).

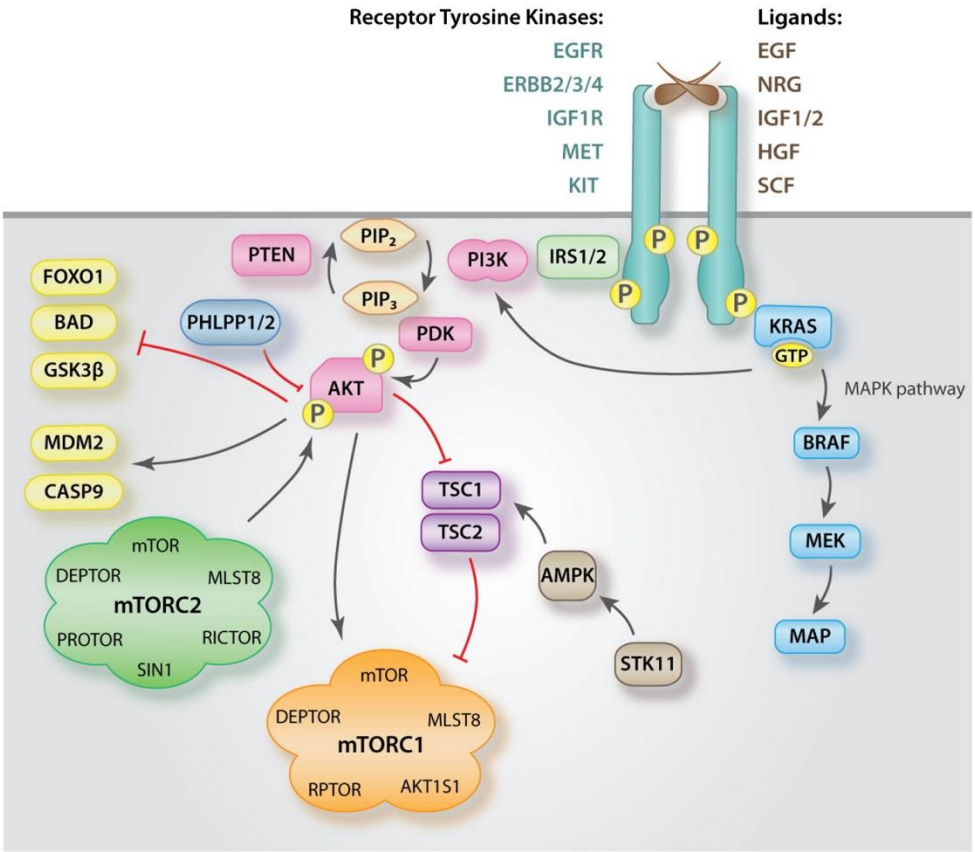

Figure 6: The signalling pathways PI3K AKT/ mTOR, on the left, and RAS/ RAF/ MEK, on the right, results in the activation of the MAP kinase like ERK, among others. In both cases activation is determined by activation of receptor tyrosine kinases due the union to its ligands. Black arrows indicate positive regulation, and red arrows indicate negative regulation. (Danielsen *et al*[44])

It should be noted that RAS/RAF/MEK/ERK and PI3K/AKT/mTOR pathways are highly interrelated forming a complex network of activations and inhibitions in the cell[48]. The importance of the PI3K pathway in CRC biological alterations, such as proliferation and cell survival, has to be highlighted, as overexpression of the gene is significantly associated with reduced relapse time of patients with stages II and III, as well as a decrease in their survival[49].

The WNT signalling route is another important pathway that favours the development of cancer cells in general, and in CRC in particular [50 51]. Among the most important genes within the WNT signalling pathway, the adenomatous polyposis coli (*APC*) is found. The determination of *APC* mutations in CRC patients is of high clinical interest because it is related to poor prognosis[52] and *APC* is mutated in 80% of CRC tumours[53]. The WNT pathway finishes with *CTNNB1* (β-catenin) activation, which induces the translocation of some transcription factors to the nucleus favouring cell proliferation [50].

On the other hand, other type of receptor, VEGFR (Vascular Endothelial Growth Factor Receptor), is related to cancer development due to its ligand, VEGF (Vascular Endothelial Growth Factor), is overexpressed in several tumours. In CRC, the overexpression of VEGF is associated with increased relapse in compare with patients who have weak expression of VEGF[54]. In addition, it should be noted that alternative splicing generates several isoforms of VEGF, generating pro-angiogenic and anti-angiogenic isoforms[55] highlighting the importance of isoforms in the development of CRC. There are different subtypes of VEGF receptors, and their functions related to angiogenesis, vasculogenesis or lymphangiogenesis[56] have to be highlighted. The activation of VEGF receptors involves the development of a complex signalling

network, where different pathways are involved, that is, activation of ERK, AKT, SRC or activation of focal adhesion kinases. Therefore, it is related not only to biological processes involved in angiogenesis, but also to proliferation, cell survival and migration[56 57].

Alterations on cyclins genes, which encoded proteins involved in the control of the cell cycle, are also found in CRC tumours, involving them in tumour malignancy, favouring proliferation and cell migration[58 59 60 61].

The cell cycle consists of a series of phases in which several genes serve as checkpoints for proper cell function. It can be divided into four phases: the mitotic phase (M), the G1, the S and the G2 phase [62 63] and it is finely regulated by phase-specific proteins comprising cyclins and cyclin-dependent kinases (CDKs)[62 64]. The progression from one phase to another goes through specific checkpoints, which are regulated by several proteins including the well-known tumour suppressors p16 (CDKN2A), p53, CDKN3 and others[65 66]. Cell proliferation pathways activated by mitogenic signals, such as those mentioned above like RAS/ RAF/ MEK/ ERK and/or PI3K/ AKT pathways end with the activation of different CDKs, favouring the entrance and the progression of the cancer cell along the cell cycle[63 61 67] (Figure 7). It is important to note that alterations in cyclins and their regulators have been associated with worse prognosis. For instance, overexpression of p16 and CDKN3 has been reported to be a poor prognostic factor in CRC[65 66].

The *TP53* gene is highly mutated in around 45% of the CRC patients[68]. *TP53* is also a marker of poor prognosis and has been used as indicative of resistance to CRC treatments[69 70].*TP53* is a tumour suppressor that inhibits the cell cycle progression in G1 phase favouring apoptosis in response to DNA damage or cellular

stress[69]. However, when *TP53* is mutated, it plays a crucial role in tumourigenesis, since it favours metastasis, cell migration and tissue invasion of tumour cells [68] [70]. As can be seen in the figure 4 *TP53* is a gene related to the last part of the sequence to generate a carcinoma from normal tissue. The loss of the chromosomal region where *TP53* is located (17q) is a common event in CRC[71]. Moreover, the mutation in *TP53* is frequently in non-hypermutated CRC[71].

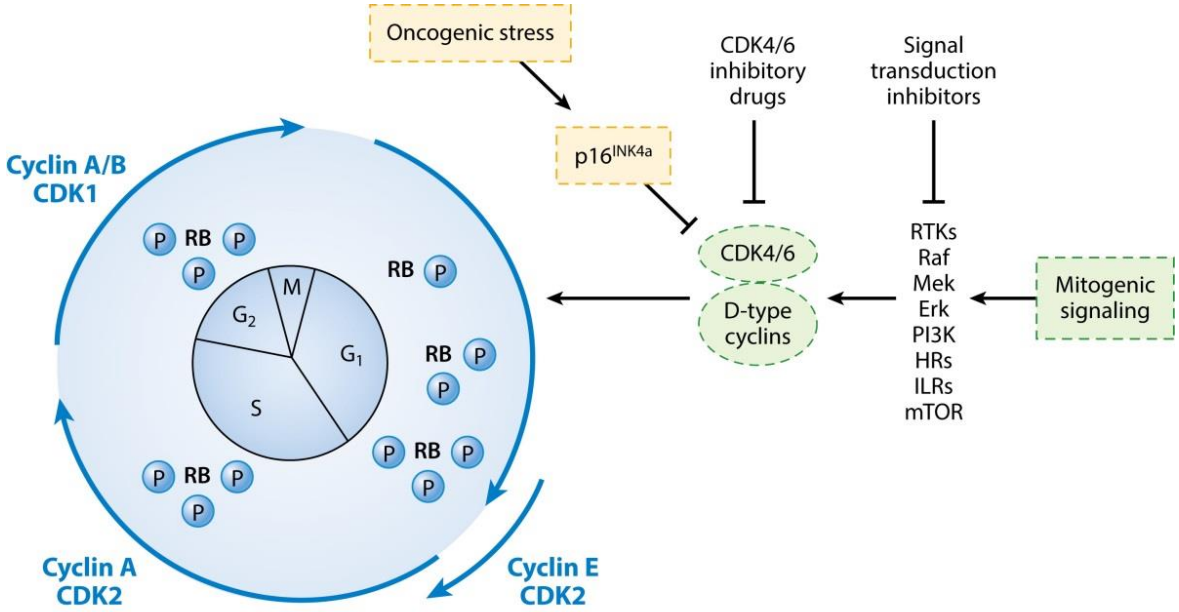

Figure 7: Activation of the cell cycle by the different genes that make the mitogenic signalling pathways related to cancer. The role of cyclins and the stages of the cell cycle are shown. (Sherr and Bartek[62])

1.5 Targeted treatments: advances in colorectal cancer

Despite the efforts, the current standard treatment for both localized and advanced CRC is cytotoxic chemotherapy [72], whose backbone is represented by the thymidylate synthase inhibitor *5-fluorouracil* (*5-FU*)[73] [74]. It can be combined with *Irinotecan* or *Oxaliplatin*[75]: the first is an inhibitor of DNA topoisomerase 1[76], while *Oxaliplatin* acts forming both inter- and intra-strand cross links in the DNA double chain[77].

Thanks to the advances in cancer cell biology, some treatments have been developed to selectively target specific molecular alterations: anti-epidermal growth factor receptor (EGFR) inhibitors and anti-angiogenic drugs. The first comprise both monoclonal antibodies (mAb) and small molecules tyrosine kinase inhibitors (TKIs) [78]. Approved mAb for advanced CRC are *Cetuximab* and *Panitumumab*, while there are no TKIs approved. Among the anti-angiogenic agents approved drugs are the VEGF inhibitor *Bevacizumab*[79] and the VEGFA and PGF (Placental Growth Factor) inhibitor *Alfibercept*[80]. All these drugs are administered in combination with chemotherapy[81] [82] [83].

While until now no biomarkers have been identified for anti-angiogenic drugs[84], the use of anti-EGFR drugs is restricted to *RAS* wild-type patients. Indeed, in RAS mutated patients the pathway is constitutively active making an anti-EGFR treatment ineffective. To date, no targeted treatments exist for RAS mutant patients apart from bevacizumab. As direct RAS inhibitors have not demonstrated efficacy, research is focusing on other downstream genes or networks such as, the combination of the anti-MEK *Trametinib* with the anti-CDK4/6 *Palbociclib* which showed promising results in preclinical models[85] [72].

BRAF mutant are a small subset of CRC patients that are characterized by a particularly poor prognosis. In this group of patients, until now the only marginal benefit was observed with the triplet combination chemotherapy of 5-FU, Oxaliplatin and Irinotecan[86]. Recently, very encouraging results have been showed with a chemo-free regimen, consisting of a combination of the anti-BRAF Encorafenib and the MEK inhibitor Binimetinib plus Cetuximab[87]. This highlights the importance of a deep understanding of molecular processes.

Trastuzumab, *Pertuzumab* and *Lapatinib* are fundamental drugs for HER2 positive breast cancer[78]. They have showed benefit in the small subgroup of HER2 amplified advanced CRC[88], being a possible treatment option (Figure 8).

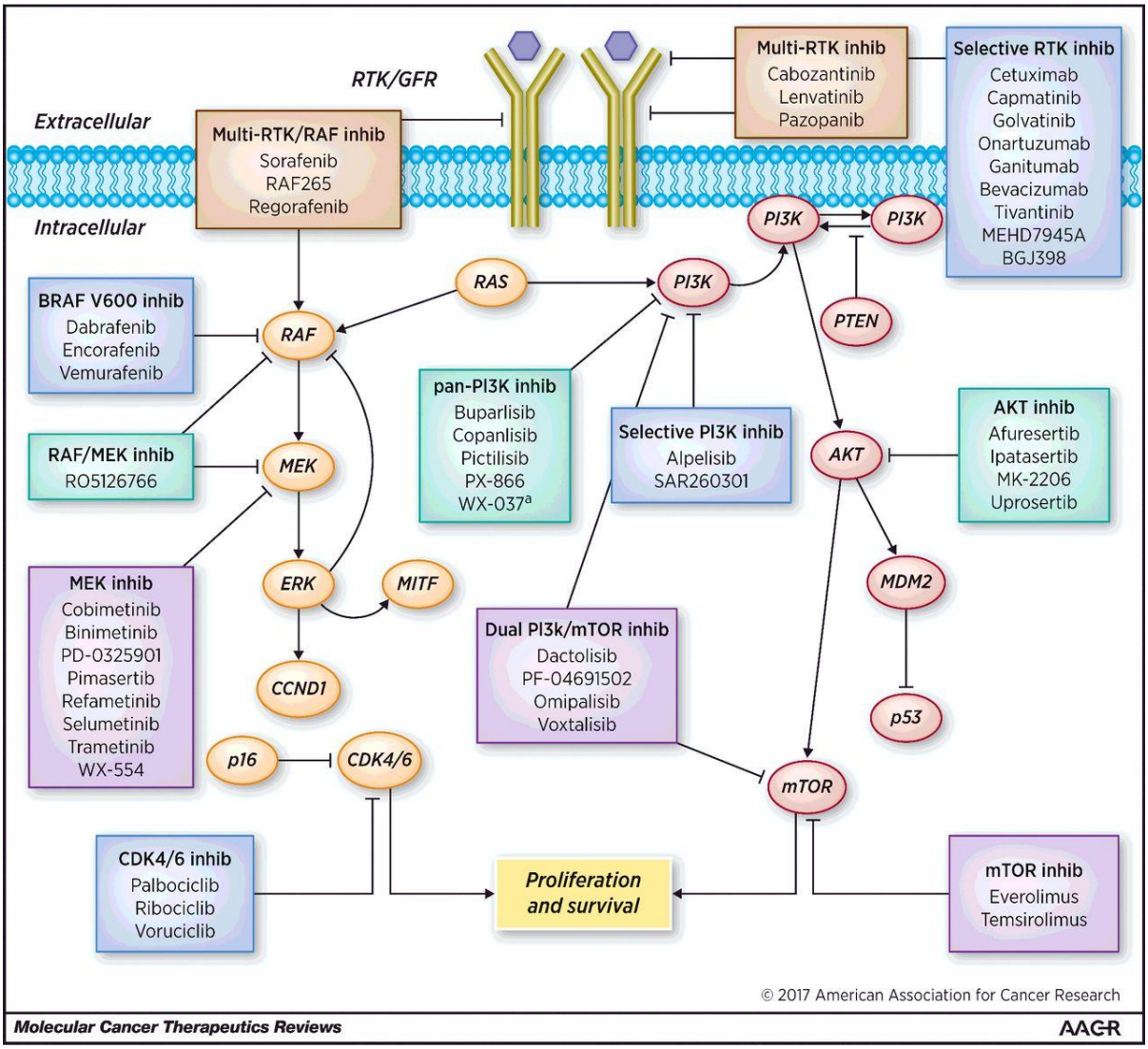

Figure 8: Overview of inhibitors in clinical development against specific targets of the PI3K / AKT / mTOR and MAPK pathways. The PI3K/AKT pathway can be activated by RAS mutants and acts as part of the cross-talk mechanism with the MAPK pathway. Both pathways can be activated by oncogenic RAS and likely serve a compensatory signaling function in cases where either pathway is inhibited (Tolcher *et al*[91])

Even RAS wild type patients do not respond in the same way to the anti-EGFR treatment and some of them present primary or acquired resistance. Described

mechanisms of resistance include mutations occurring in genes that are downstream in the EGFR or in other pathways, such as *PI3KCA, AKT, mTOR, RAF, MEK* or *ERK*; this means that, despite the inhibition of the receptor, the route will not be inhibited and the proliferation pathways would remain constitutively active. In these circumstances, directly targeting downstream effectors would be much more effective. Inhibitors that act, for instance, on mTOR (*Everolimus)*[89] or AKT (*Afuresertib)*[90] among others have been studied (Figure 8).

1.6 Tumour microenvironment in colorectal cancer

The constant cell replication to form a tumour mass in intestinal epithelium, must be supported by different factors and adjacent cells that provide nutrients and stabilization for the correct development of the tumour. The tumour microenvironment is formed by fibroblasts, myofibroblasts, neuroendocrine and adipose cells and immune and inflammatory cells, as well as an extensive blood and lymphatic network[92].

One of the main characteristics of the tumour cells to be highlighted is the capacity of immune evasion[1]. Under normal circumstances, cytotoxic or $CD8^+$ lymphocytes and helper or $CD4^+$ T lymphocytes inhibit aberrant cells growth involving production of interferon-γ and cytokines[93], as well as macrophages and fibroblasts. In contrast, tumour-associated macrophages (TAM) and cytokines such as IL-6, IL-1β, IL-23 or TNF promote tumour cell development of cancer cells[94].

There is a double role of the immune system in relation to tumour development. This double relationship occurs because cell families surrounding the tumour are "transformed" by differentiating factor signals produced by the carcinogenic cells[95].

Thus, the macrophages located around the tumour tissue, and transformed into TAM, secrete cytokines and pro-tumourigenic growth factors[96]. Likewise, regulatory T lymphocytes transform several immunological response pathways, favouring the inhibition of B and T lymphocyte response and antigen presentation by dendritic cells[95 96]. By contrast, in normal conditions the regulatory T cells favour lymphocytic homeostasis.

In parallel with transformation of macrophages to TAM, the fibroblasts undergo a transformation to cancer-associated fibroblasts (CAFs) due to factors secreted by the tumour[96 97]. The CAFs have a capital role in angiogenesis and, consequently, in metastasis, since formation of blood vessels allows tumour irrigation which, together with the role of CAFs in extracellular matrix secretion, favours invasion and migration[95 97].

Under such conditions, tumour cells can enter the blood vessels thanks to the help of the TAM, and once inside the blood vessels, the tumour cells are surrounded and protected against the immune system by the platelet action[95]. In this way, thanks to the microenvironment that surrounds the tumour, the cancer cells avoid the immune system and can move and adhere to another tissue generating metastasis (Figure 9).

It should be mentioned that there is an increasing number of studies on cancer molecular biology, especially in CRC, which focus their attention not only on the tumour mass but also on the peritumoural areas. The study of the so-called "*healthy tumour borders*" seems to be essential to understand the genetic and molecular mechanisms that favour the dispersion of the primary tumour. Although anatomically and macroscopically the tissue adjacent to the tumour is usually called healthy tissue, recent studies show that those areas, exhibit some genetic transition from normal to

tumour tissue, with large molecular differences between healthy tissue and peritumoural tissue[98]. This fact has been demonstrated in CRC patients, determining a crossover of protein secretion signals and receptors activation between the tumour and peritumoural colon tissue[99].

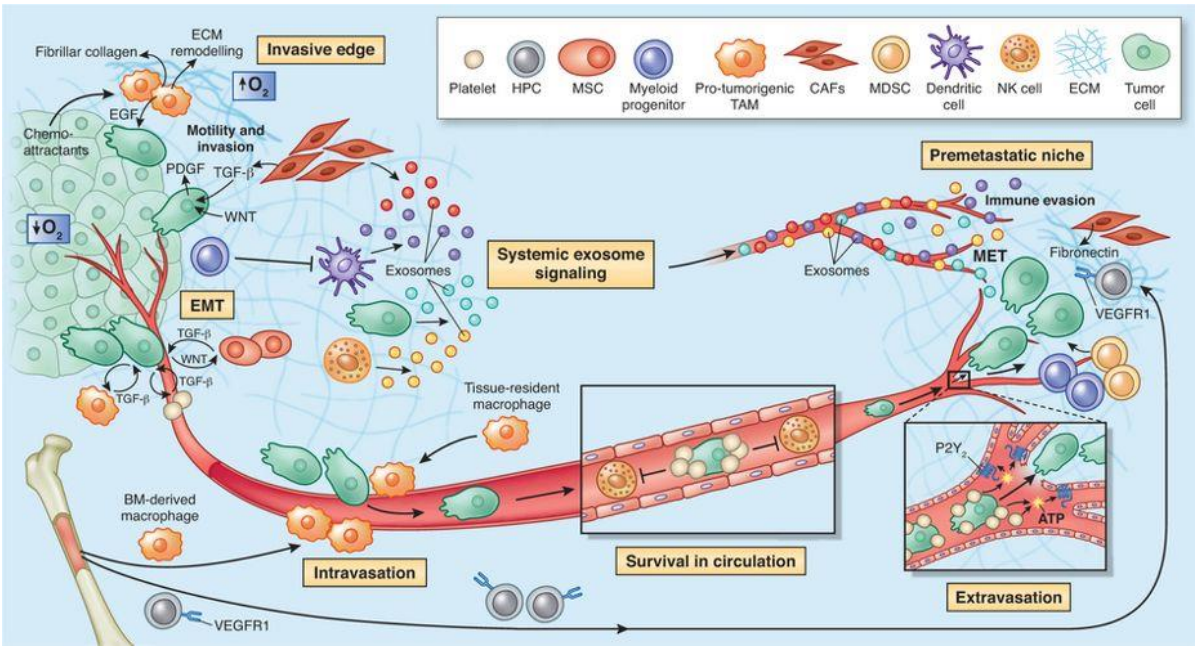

Figure 9: Activation process of EMT favoured by the components of the immune system that surround the tumour cells. HPC: hematopoietic progenitor cells. MSC: mesenchymal stem cells. MDSC: myeloid-derived suppressor cell. NK: natural killer cells. (Quail and Joyce[95])

1.7 Epithelial-mesenchymal transition and its role in colorectal cancer

Epithelial-mesenchymal transition (EMT) is a complicated cellular process that occurs during normal embryonic development to allow epithelial cells more amenable to move and escape from the structural constraints of tissue architecture. In this process epithelial cells acquire a mesenchymal phenotype. Recently, some events of physiological EMT, specially the loss of epithelial markers and the gain of mesenchymal markers, have been described in pathological states, including epithelial cancers such as CRC. In cancer cells, EMT is abnormally regulated by

microenvironment, including growth factors cytokines and also intratumoral stresses like hypoxia and is associated with an increase in the invasive and migratory capacity of tumour cells favouring progression and metastasis[100].

As it is mentioned normal epithelial cells have some characteristics, such as non-mobility and non-invasive capacity, and possess some molecular markers of this state, such as the expression of E-cadherin and Claudin among others[101]. The *CDH1* (E-cadherin) is defined as a tumour suppressor gene, which favours cell-cell contact. The loss of its activity facilitates cancer invasion and migration[102]. A similar role is described for Claudin, a transmembrane protein whose action lies in the maintenances of cell-cell junction[103].

As it is mentioned tumour cells acquire mesenchymal phenotype and this fact favour mobility and invasive capacity. Mesenchymal cells are also characterise by specific markers such as N-cadherin, Vimentin, Snail or Fibronectin[101]. N-cadherin overexpression is highly related to an increase in mobility and invasive capacity[104]. On the other hand, Vimentin plays an important role in the organization of the cytoskeleton which favours focal adhesion[105] and, also, its overexpression in CRC is indicative of poor prognosis[106]. Snail, a zinc-finger transcriptional factor, induces the transition to mesenchymal cell inhibiting the transcription of E-cadherin and Claudin[107] having a role in tumour development, metastasis and drug resistance[108]. Also, Slug, another zinc-finger transcriptional factor encoded by *SNAI2*, has similar functions to Snail and has been demonstrated its overexpression with the CRC progression[108 107]. Finally, fibronectin, a glycoprotein of the extracellular matrix, is crucial in cell adhesion, migration and metastasis. Moreover, it has been shown that overexpression of fibronectin correlates positively with more advanced stages in patients with CRC[106 95].

1.8 Epigenetics of colorectal cancer

During the last years, several studies have shown the implication of different epigenetic mechanisms in the development of CRC. Epigenetics could be defined as the study of changes in gene function, caused by mechanisms other than DNA sequences that have even been shown to be heritable[109].

Epigenetic mechanisms range from the modification of histones, through processes of methylation or acetylation and the methylation of DNA, to the interferences generated by non-coding RNA, such as microRNAs, lncRNAs or siRNAs (Figure 10)

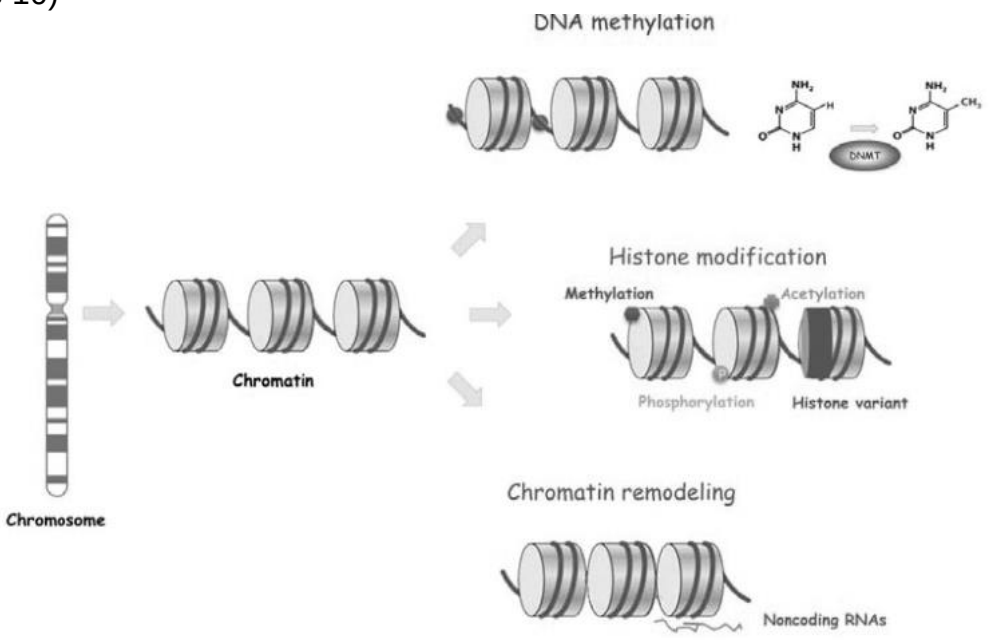

Figure 10: Changes in gene expression due to variations in chromatin structure, not in direct modifications in the DNA sequence. Such chromatin variations may be due to DNA methylations (above), histone modification by methylation, acetylation and phosphorylation (in the middle), or chromatin remodelling due to nucleosome or non-coding RNA occupancy (below) (Choi and Lee[109]).

DNA methylation is a very important area of study in the development of cancer, and specifically in the CRC [110]. The CRC cells are characterized by having a global DNA hypomethylation resulting in an instability in the cell genome involving the

aberrant activation of oncogenes[110]. In the CRC cells, hypermethylation of certain CpG islands is also found. When located in the promoters of some tumour suppressor genes, it triggers different carcinogenic processes[111]. DNA methylation is carried out by DNA methyltransferases (DNMTs), which add methyl groups to cytosines to CG dinucleotides[112]. Some of the CpGs are accumulated, forming CpG islands, in gene promoters, where transcription factors and RNA polymerase must bind to carry out the correct transcription of the gene[113]. If the CpGs are highly methylated, methyl binding domain (MBD) proteins are recruited and physically prevents transcription factors and RNA polymerase binding to the promoter[110]. MBD proteins also recruit histone deacetylases, leading to the formation of a compact chromatin structure that hinders transcription (see below).

The structure of the chromatin as well as its compaction, in turn, is essential for the normal functioning of the cells. Chromatin structural changes generate epigenetic dysregulations, which could be involved in the development of cancer[114]. The chromatin consists fundamentally of the nucleosome, which has 147 base pairs of DNA wrapped around an octamer of core histone proteins, H2A, H2B, H3, and H4[114].

Alterations in the modification of histones could be an epigenetic process of great relevance for CRC development that is largely unexplored. The different modifications carried out on histones imply differences in chromatin folding, modulating the expression of the gene involved[115]. The histones have an N-terminal end that protrudes from the nucleosome where most of the modifications, for instance acetylation[116 117], methylation[115 118 119] and phosphorylation occur[120 117] (Figure 11). The enzymes responsible for acetylation are called histone acetyltransferases (HAT),

meanwhile the enzymes that remove acetyl group from the histones are the histone deacetylases (HDAC)[121] [122].

Several studies have related different histone modifications with CRC development by modifying specific gene expression and triggering pro-tumourigenic pathways[123]. Even it has been proposed that these posttranslational modifications of histones in circulating nucleosomes could be used as biomarkers in CRC [124]. The modification of histones by acetylation is fundamentally related to transcriptional activity, because it leads to a structure of chromatin more accessible. On the other hand, deacetylations contribute to the compaction of chromatin[125] [126]. Interestingly, a clear dependence between the age of CRC patients and the modifications H3K9ac and H3K27me3 has been identified[127]. Apart from being considered as prognostic value of the disease, it also makes clear differences between the biology of the CRC between young and adult patients [10]. Figure 11 shows the different modifications that histone tails can undergo, changing the pattern of genetic expression.

The microRNAs are other epigenetic regulators of the CRC. The microRNAs are small molecular RNAs of around 20 to 25 ribonucleotides, which interfere in gene expression through post-transcriptional regulation. The regulation by microRNAs of different cellular signalling pathways, as well as their contribution in CRC, favouring its progression and metastasis, has been demonstrated in different investigations[129] [130].

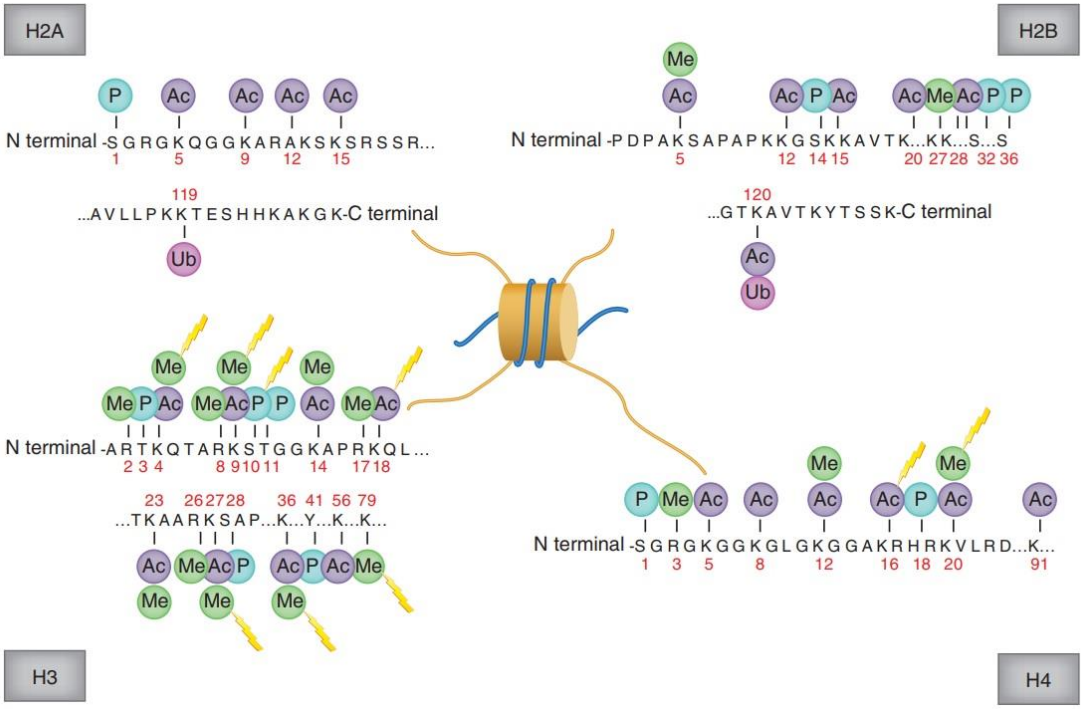

Figure 11: Histone modifications patterns. The different modifications can be observed on the N-terminal histone tails, also on the C-terminals. The modifications of the histone tails pattern, determines in a large way the chromatin condensation, and therefore, the variations in the gene expression. It can be seen in the figure the different types of modification such as phosphorylation (P), Acetylation (Ac), Methylation (Me) and Ubiquitination (Ub) in the four types of histones, H2A, H2B, H3 and H4. (Rodríguez-Paredes and Estellés[128]).

In addition to regulating different processes related to oncogenesis, microRNAs act as mediators in communication not only from cell to cell, but also between the primary tumour and its metastasis by exosomes[131]. The microRNAs show an imperfect recognition on their target, aided by DICER and Argonaut proteins, favouring gene silencing due to the lack of translation of mRNAs[132]. The case of siRNAs is different since, by a perfect recognition on their target, favours mRNAs fragmentation[133].

On the other hand, the long non-coding RNA (lncRNA) act as scaffolds to attract and anchor complex chromatin remodelling sites [134], playing a role in the recruiting of factors to specific target sequences. The participation of lncRNAs as key gene

regulation in CRC has been described in the screening, diagnosis and prognosis of the CRC[135], as well as in changes in migration, invasion and worse survival of CRC patients[136].

1.9 Alternative splicing in colorectal cancer

Alternative splicing is a process that play a central regulatory role in the control of gene expression. Through this process a single gene usually gives rise to several mRNA isoforms. The different isoforms may possess different roles, hence the importance of alternative splicing in regulating gene function. The alternative splicing can generate several common splicing patterns, including alternative 5´splice site, exon skipping, intron retention, alternative 3´splice site and mutually exclusive exons[137] (Figure 12). The mechanisms involved in the selection of one or another isoform are complex and they are not yet completely understood, although it is known that the chromatin structure and its epigenetic modifications play an important role[138]. Hundreds of proteins, like snRNPs and hnRNPs (small nuclear and heterogeneous ribonucleoproteins), are related to the splicing and act as regulators of the process[139]. After that, the hnRNPs play an important role in the stabilization of the mature mRNA. It is important to highlight that the hnRNPs have the capacity to bind to 3´and 5´ UTR of mRNAs, so they can control translational repression or enhancement[140].

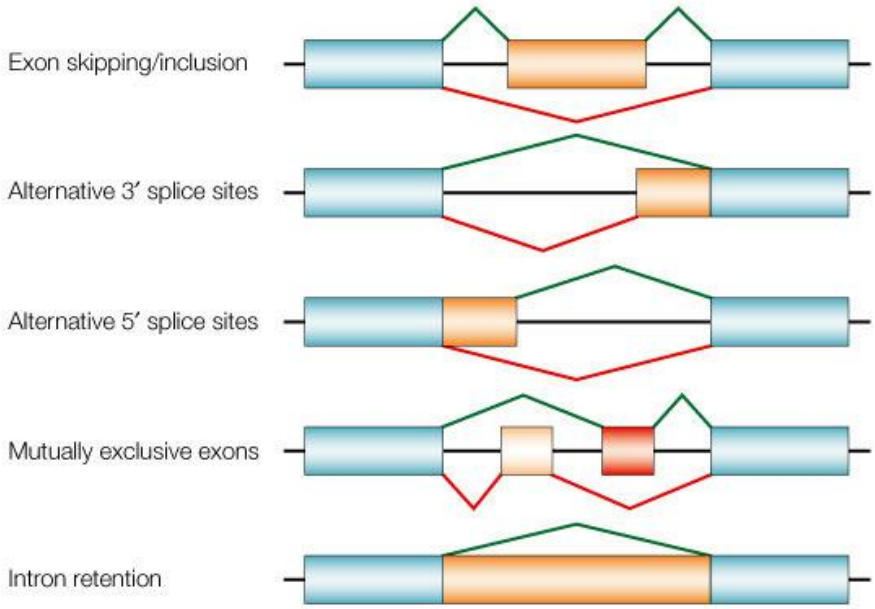

Figure 12: Schematic of constitutive and alternative splicing events (red and green joining lines). The blue boxes represent constitutive exons, and yellow and red boxes represents the alternatively spliced exons. Introns are represented as black lines between the boxes (Cartegni *et al*[141]).

In the case of CRC the splicing of many genes has been shown to differ between normal mucosa and tumours[137]. In some instances, alternative splicing is involved in the function of the regulatory pathways determining CRC progression[142]. Important genes related to previously described pathways such as the WNT pathway undergo alternative splicing processes. It has been described that WNT5A has two isoforms, the so-called L (long) and S (short). In the case of isoform L, it has been described as a suppressor of cell proliferation in CRC, while S has been described as a promoter of proliferation having an important oncogenic role in CRC[143]. A similar example occurs with the Cyclin-D1 (*CCND1*), *CCND1* acts as a regulator to the CDK4-CDK6 complex, which promotes the cell division, favouring the G1-S transition. Recent studies have demonstrated the existence of two isoforms of *CCND1*, *CD1a* and *CD1b*. *CD1b* cannot be translocated to the cytoplasm, so it is retained in the nucleus,

favouring the CDK4-6 binding, on the other hand, *CD1a* translocates to the cytoplasm, so there is no data on its activity. In this way a pathogenic isoform related to the development of cancer is observed, and another isoform unknown in its function[144].

On the other hand, it is important to highlight the role of the alterative splicing on the KRAS gene[145]. *KRAS* has 4 alternative splicing isoforms, two of which are translated into proteins, the *KRAS*-4A isoform and the *KRAS*-4B isoform. The alternative splicing process on *KRAS* is performed by combining its 7 exons and 6 intronic regions, its exons being translatable from 1 to 4A or 4B, according to the isoform [146 147]. Both the 4A and 4B isoforms bind tightly to the inner sheet of the plasma membrane through electrostatic interactions due to a tract containing 7 lysines [148]. The route by which they arrive at their final destination differs between the isoforms 4A and 4B. *KRAS-4A* is directed to its location via the Golgi system, while *KRAS-4B* goes to the plasma membrane through the endoplasmic reticulum [149]. The ratio of the two isoforms are altered in cancer, but while *KRAS-4A* is considered pro-apoptotic, *KRAS-4B* is antiapoptotic, so isoform 4B has a greater oncogenic role[146 150].

1.10 KRAS mutational status and its phenotypical consequences in colorectal cancer

The importance of the *KRAS* study lies in its high mutational rate in various tumours, especially in CRC, causing a constitutive activation of the proliferation pathway RAS / RAF / MEK / ERK[151]. Moreover, as it is mentioned above it has been demonstrated that patients with KRAS mutations may not benefit from anti-EGFR therapy, a major step forward in the treatment of metastatic colorectal cancer.

The most relevant, clinically important mutations in *KRAS* are usually found in codons 12 and 13 of exon 2, other mutations only have minor interest in CRC[152 153]. Mutations in *KRAS*, especially in codon 12, are related to a worse patient prognosis[154]. Mutations G12V and G12A are especially relevant, since progression-free survival in these patients with metastatic CRC is 6.6 months compared to 11.6 months in patients with other mutations in *KRAS*[155]. Another interesting detail is that a significant increase of mutations in codon 13 is found in the right side of the colon with respect to the left one[154]. Also, it has been described that G13D *KRAS* mutations mediate the silencing of tumour suppressor genes by hypermethylation of the CpG islands of their promoters[156].

In line with these epigenetic alterations associated with *KRAS* mutated in CRC, previous published results of our group demonstrated that global acetylome of CRC cells was dependent on *KRAS* mutation status[157]. So that, several hnRNP family members showed higher acetylation associated to the presence of *KRAS* G13D mutation in CRC isogenic cell lines. In this study it was also observed how the acetylation of hnRNPs remained unchanged, despite the treatment with EGF growth factor, in CRC cell lines with the *KRAS* G13D mutation. In contrast, cells that contained another mutation *KRAS* named A146T the EGF treatment did favour the increase in acetylation of the hnRNPs[157]. Also, the status level of hnRNPs acetylation is higher in cells with G13D mutation in *KRAS*, regardless of treatment.

In view of these results a deep study to analyse the functional consequences of acetylation changes in splicing factors in the context of G13D *KRAS* mutated cells was performed. An *in silico* analysis of RNA-seq studies, published by our group in 2016[158], identify a list of genes with a distribution of alternative splicing isoforms

altered in CRC cell lines carrying the G13D *KRAS* mutation when compared with wild-type *KRAS* cell lines. Among these genes two of them, *ZNF518B* and *EPDR1* genes, were further validated experimentally by quantitative real-time PCR. Expression of these genes was negligible in the *KRAS* wild-type cell lines, Caco2, RKO and SW48 compare with cell lines that harbour mutant G13D *KRAS* gene, HCT116, DLD1. Also the ratio of alternative splicing isoforms was significantly different in isogenic cell lines D-Mut1, derived from DLD1, in which the wild-type allele had been knocked out, so only contain one G13D mutated allele.

Above mentioned data of our group suggested a relationship between *KRAS* G13D mutation and the expression of total *EPDR1* and *ZNF518B* genes and their isoforms marking the beginning of the work presented in these thesis.

KRAS mutations in CRC are common in human and as it has been described above they cause resistance to anti-EGFR drugs used for the treatment of metastatic CRC. There are still many unsolved questions on the molecular mechanism downstream of *KRAS* that could be involved in this resistance so any effort to increase our knowledge in this area would be of great interest. In these sense, the purpose of the work presented here was to perform a deep study of both *EPDR1* and *ZNF518B* genes that could enhance understanding not only on the molecular mechanisms involved in the resistance but also on the key factors involved in the progression of CRC.

1.11 Previous data of *EPDR1* gene

The *EPDR1* gene encodes a type II transmembrane protein related to ependymins, a family of glycoproteins first described as related to learning processes in teleost fishes in 1979[159]. *EPDR1* is a gene located on chromosome 7, where it is partially overlapping with the *NME8* and *GPR141* genes. The gene has 6 exons and 5 introns whose different combinations provide 4 isoforms of the gene. The 4 isoforms contain ORFs putatively coding for protein, the canonical one being that corresponding to the translation of the isoform 1 mRNA (Figure 11). As it described above, our group described that the isoform 2 of the *EPDR1* gene is expressed to a low level and it depends in CRC cell lines on the G13D mutation in *KRAS*[158]. The expression of isoforms 3 and 4 is only residual.

It is interesting to note that isoform 2 lacks the topogenic signal for membrane localization[158], while isoform 1 codes for the transmembrane functional protein (Figure 13). Due to historical reasons, the *EPDR1* gene is also designated as *UCC1* (upregulated in colon cancer 1) or *MERP1* (mammalian ependymin-related protein 1).

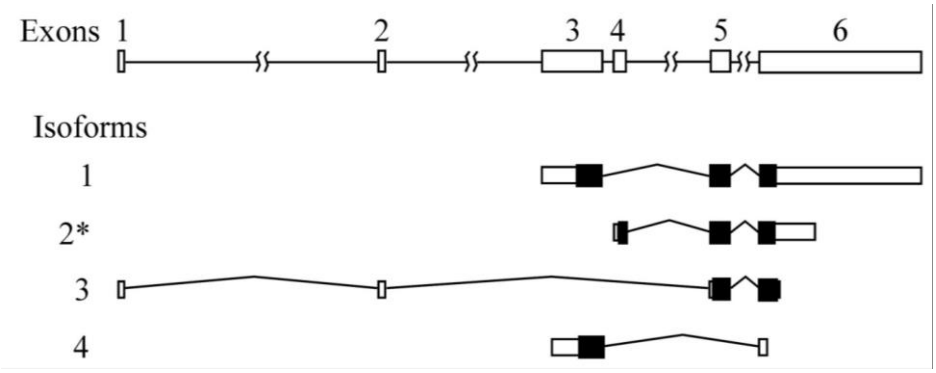

Figure 13: Exon-Intron organization of *EPDR1* gene and its different isoforms generated by alternative splicing. The translatable areas corresponding to isoforms can be observed in black, on the other hand the non-translated exons are depicted as open boxes. Introns are represented by the lines that join the boxes. The isoform reported as dependent on *KRAS* G13D mutation is marked by an asterisk. (Riffo-Campos *et al*[158])

The study of a transmembrane protein always generates a high interest, because they are good targets for the development of drugs, as their ectodomains may be easily accessible to them. They are usually at the beginning of signal transduction to the interior of the cell, activating different genetic cascades and also can have different functions such as the transport of ions and molecules to the interior of the cell. Moreover, they may participate in cell-substrate or cell-cell interactions[160].

The structure of the *EPDR1* protein is characterized by possessing a C-terminal domain containing glycosylated sites of Ca^{+2} binding and an intracellular N-terminal domain[161 162]. Calcium binding produces conformational changes in ependymins, and this modification is important for interaction with the extracellular matrix, specifically with collagen. The structural data of isoform 1 are compatible with a clear example of transmembrane protein[163] although, it has also been shown that *EPDR1* has an intracellular location, specifically associated with lysosomes[164]. It remains to be determined whether the lysosomal protein corresponds to isoform 2, which lacks the signal peptide, or to a truncated form of isoform 1.

Recently, the structure of EPDR1 has been defined by crystallization analysis[165]. Structural studies related to the EPDR1 protein indicate, at a monomeric level, that the protein is made up of 11 antiparallel leaf-β and 2 α-helix chains (Figure 14).

Contrarily to what happens with *EPDR1* gene, the expression of ependymins family genes has been widely studied. A high expression of ependymin has been observed in organs such as the brain[164]. Classically, ependymins have been assigned to a direct role in the neurochemical involvement of memory and neuronal regeneration, so it is not surprising that the first studies were performed in the brain[166].

The increase in the expression of ependymins during neuroplasticity has been demonstrated, with crucial functions in the establishment of intercellular contact during neural and optic nerve regeneration in different animal species[167]. These previously discussed functions allowed to conjecture a role for ependymins related to the extracellular matrix, possibly as a possible new class of anti-adhesive molecule [166] [167].

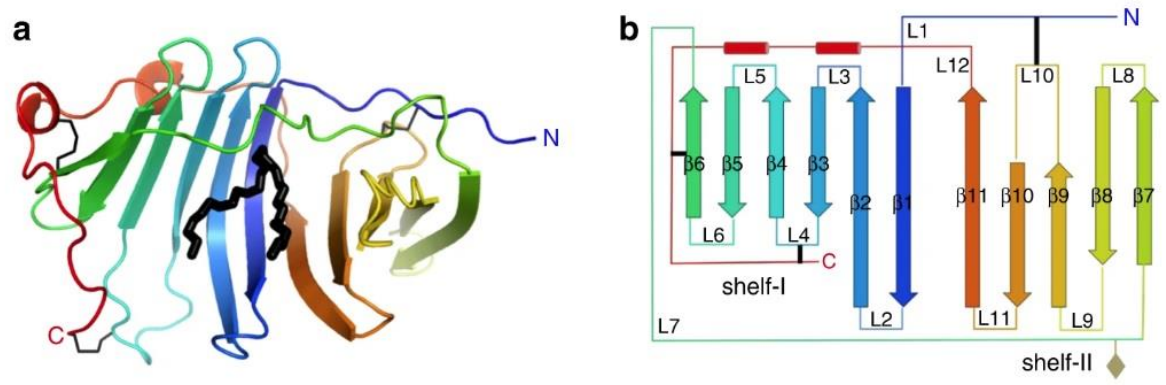

Figure 14: Structure of EPDR1 homodimer protein. Graphical structure of the EPDR1 protein in Figure a, to the left (3D structure); in Figure b, to the right (structure in plane). Observe the black junction lines, representing the disulfide bridges. The glycosylation site is represented in Figure b, as a diamond, located in shelf-II (L7). You can see the 11 β sheet and the 2 α helix in L12, before the c-terminus (Wei *et al*[165]).

In mammals, EPDR1 has been detected and studied in organs such as the adrenal glands and prostate. Specifically, it was determined that *EPDR1*, among many other genes, presented changes in single nucleotide polymorphisms (SNP) that could predict damages of radiotherapy treatment in patients with prostate cancer[168]. In a transcriptomic analysis of the overexpression effect of the protein osteopontin, it was observed a relevant change of EPDR1 gene expression in prostate and ovary cell lines. Interestingly, genes found to be altered together with EPDR1 have common functional profile as they are directly or indirectly modulate angiogenesis, vascular permeability, lymphatic metastasis or tumour-stroma interactions[169].

Another property of *EPDR1* is that it is significantly overexpressed in $CD34^{+}$ hematopoietic cells when compared to $CD34^{-}$[163]. This is interesting because CD34 is a marker in various types of cancer, including CRC[170]. CD34 is expressed in stem cells precursor of endothelial cells of small vessels, it should be noted that a significant correlation of the density of CD34 expression in intratumoral microvessels, with the stage in CRC has been observed[171].

An interesting publication by Staats K.A *et al* [161] demonstrated that *EPDR1* was directly related to the development of Dupuytren's disease. This pathology is characterized by a progressive and irreversible proliferation of fibroblasts affecting the palmar fascia[161]. It should be noted that in the study of patients with Dupuytren, they have detected *EPDR1* expression in both healthy patients and patients with the pathology, but there is a SNP that is only found in patients who suffer from the disease. The authors also demonstrated that a decrease in the expression of *EPDR1* is directly related to the decrease in collagen contraction capacity[161].

Finally, in relation with CRC there is one publication of Nimmrich *et al*[162] that demonstrate for the first time, using paired samples of 3 patients, that *EPDR1* is overexpressed in tumour samples of CRC with respect to the healthy sample adjacent to the tumour[162].

As it was mentioned above, our group published[158] that the expression of whole EPDR1 gene and its isoform 2, produced by alternative splicing, showed clear differences among CRC cell lines being dependent on *KRAS* G13D mutation. Canonical isoform 1 of *EPDR1* encodes a type II transmembrane protein related to piscine ependymines, proteins that have been involved in cell adhesion. The sequence containing the topogenic signal for membrane localization is absent in isoform 2, and

this may imply a different functional role for the encoded protein. A limited previous study with only 3 CRC patient samples suggest that EPDR1 could be overexpressed in human CRC so this fact and in view of all the above comments, it seems very interesting studying the possible implication of EPDR1 gene and its isoforms in the oncogenic processes of CRC.

1.12 Previous data of *ZNF518B* gene

Human *ZNF518B* gene is located in chromosome 4 and is transcribed from the reverse strand. The gene encodes a zinc-finger protein but the knowledge on the role of this gene is very limited. According to the ENSEMBL database, *ZNF518B* gene is made up of 5 exons and 4 introns, and alternative splicing may result in a total of 6 isoforms of the gene. The isoform 1 codes for the canonical protein of 1074 residues while isoform 2 putatively produces a truncated form of the protein by the N-terminal domain, so the latter does not possess the zinc fingers of ZNF518B isoform1 (Figure 15).

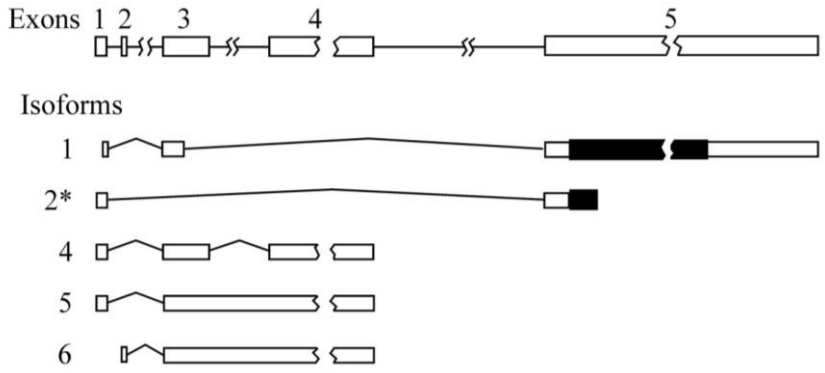

Figure 15: Exon-Intron organization of *ZNF518B* gene and its different isoforms generated by alternative splicing. The translatable areas corresponding to isoforms 1 and 2 can be observed in black; on the other hand the non-translated exons are depicted as open boxes. Introns are represented by the lines that join the boxes. The isoform reported as dependent on KRAS G13D mutation is marked by an asterisk. (Riffo-Campos *et al*[158])

Proteins with zinc finger domains, like ZNF518B, have different functions at the cellular level, including DNA recognition, RNA packaging, transcriptional activation, protein assembly and lipid-binding capacity [172]. The genes that code for the zinc fingers represent 3% of the genes in the human genome[173]. Structurally, the zinc fingers domains contain one or two zinc ions coordinate with histidine and cysteine aminoacid residues[174]. The interaction between zinc finger-containing proteins and DNA was cleared after the crystallographic studies of Pavletich and Pabo[175] indicating that the zinc fingers interact with the major groove of DNA[176].

There are several examples of the involvement of zinc-finger transcription factors in CRC, such as the GATA transcriptional factor. The GATA protein is formed by two zinc fingers and has an important role in the differentiation and homeostasis in the intestinal epithelium[177]. Several studies have shown that GATA favours hypermethylation and transcriptional silencing in CRC[178]. KLF4 is another crucial zinc finger transcriptional factor in the differentiation of the calceiform cells of the intestinal crypts[179]. It has been observed that the expression of KLF4 in healthy tissue is higher than in tumour in CRC, in addition its under-expression in both healthy and tumour tissue has been associated with low survival[180]. Regarding ZNF281, its involvement in the development of metastases in CRC through its regulation of EMT, has been demonstrated[181]. ZNF281 acts by inducing the expression of *SNAI1*, *ZEB1* and *VIM*, so these studies conclude that ZNF281 is a zinc finger transcriptional factor with oncogenic functions[181 182]. Another example is that of *ZKSCAN3*, whose mechanism in the tumourigenesis consists in favouring cell proliferation, migration and angiogenesis[183 184]. Finally *ZNF304*, whose concentration is regulated by the *KRAS* G13D mutation, acts as a tumour suppressor gene facilitating the recruitment of

DNMT1 (DNA Methyltransferase 1), an enzyme that catalyses the transfer of methyl groups to specific CpG dinucleotides in DNA[156 184].

As mentioned above, little is known about the function of *ZNF518B*. In 2015 Tian-Bo Jin *et al*., after a genome-wide analysis, described *ZNF518B* as a diagnostic and prognostic marker in patients with gout disease.[185]. In this same line of research, other studies demonstrated the correlation of certain polymorphisms of *ZNF518B* with the increase in gout in the Tibetan population, in addition to linking them with high glucose levels in the patients studied[186]. In a further study of glucose metabolism, researchers showed that the silencing of *ZNF518B* on the β cells of the islets of Langerhans produces alterations in the capacity of insulin secretion[187].

It is worth highlighting the work published by Maier *et al.* 2015*,* where it was demonstrated that ZNF518B interacts with histone methyltransferase G9a and regulates its activity[188]. The deregulation of this enzyme has been involved in cancer. Hypermethylation of lysine 9 of histone 3 ($H3K9me^3$) catalysed by histone methyltransferase G9a, has been detected in many types of cancer and associated with a poor prognosis[189]. H3K9 methylation is recognized by the heterochromatin 1 protein, favouring the packaging of chromatin and therefore transcriptional repression. G9a undergoes hyperactivation under hypoxic conditions, allowing the repression of genes in response to hypoxia, a condition very frequent in solid tumours, so G9a could function in these conditions as an enhancer of cell survival, proliferation and metastasis[190].

As mentioned above, Isoform 2 of *ZNF518B* does not possess the zinc finger domains present in the canonical isoform 1. Nevertheless, in view of the results of Maier *et al*. 2015, it may be possible that isoform 2 would be still able to interact with

G9A playing a role by competing with the whole isoform 1 [188]. The role of *ZNF518B*, a poorly studied gene, with an expression profile dependent on *KRAS* mutated status, and the relationship between G9a and *ZNF518B* isoforms in the onset of colorectal malignancies seems a very interesting questions to further study.

OBJECTIVES

Objectives

In the frame of the antecedents described above, the following objectives were planned, in an attempt to solve some of the questions concerning the possible role of *EPDR1* and *ZNF518B* in CRC

1. To quantitatively determine the expression of whole *EPDR1* and *ZNF518B* genes, as well as of their major and significant isoforms in samples of CRC patients in comparison with healthy tissue samples.

2. To correlate the expression level of *EPDR1* with clinico-pathological variables in a prospective cohort of patients of Hospital Clínico de Valencia.

3. To evaluate the relative ratio of the two main alternative splicing isoforms of *EPDR1* and *ZNF518B* genes in patient samples. To determine whether a correlation with clinico-biological and pathological parameters exists.

4. To study the role of *EPDR1* and *ZNF518B* genes in CRC cell lines by analysing the phenotypic consequences of the knocking-down and overexpression of the genes.

5. To determine the genetic and epigenetic mechanisms involved in the regulation of *EPDR1* and *ZNF518B* gene expression in CRC.

6. To check whether the expression of *EPDR1* gene allows relating it to any of the four CMS subtypes of the molecular classification of CRC.

7. To deepen the knowledge of the potential functions of *ZNF518B* in CRC by exploring the nature of the possible targets of ZNF518B as a transcription factor in a genome-wide transcriptomic assay.

MATERIALS AND METHODS

Materials and methods

3.1 Cell lines, cell culture and reagents

During the development of these studies, different human CRC cell lines have been used, such as SW48 and RKO (Horizon Discovery, Waterbeach, UK), HCT116 (ATCC CCL-247), DLD1 (ATCC CCL-221) and its derived isogenic line D-Mut1 (courtesy from Dr. B. Vogelstein, Johns Hopkins School of Medicine, Baltimore). The mutational state of the cell lines is described in Table 2. DLD1, D-Mut1 and HCT116 cell lines were grown in McCoy´s media (Sigma Aldrich) and RKO and SW48 cell lines were grown in RPMI media (Sigma Aldrich). In all cases, the media were supplemented with 10% foetal bovine serum (GE Healcare, cat: SV30160.03), 1% L-glutamine (Biowest, cat: X0551-100) and 1% penicillin/streptomycin (Biowest, cat: L0018-100). The cells were cultured at 37ºC in a humid atmosphere containing 5% CO_2.

To guarantee the identity of the different cell lines along this study a Short Tandem Repeat (STR) DNA profiling has been performed by the biotechnological company of genetic analysis Bioidentity (Elche, Alicante). Genomic DNA was extracted and analysed confirming the identity of cell lines according to standard and international authentication parameters (ANSI/ATCC).

For all the experiments, cell lines were used in early passages, less than seven. For cell maintenance, cultures were prepared from 1:6 dilutions when confluence was near the 90%. To this purpose, a washing with phosphate-buffered saline (PBS) was carried out, cells were subsequently raised with 0,25% trypsin-EDTA 1x (Gibco), and seeded in their corresponding medium.

For cryopreservation, cells were suspended in FBS with 10% of dimethylsulfoxide (DMSO) (Sigma Aldrich). Cells were stored in cryovials and placed at -80ºC for 1 day in a cool-container before being placed in a liquid nitrogen tank for long term storage.

Table 2: Relevant genotype of cell lines used in the present study.

Cell lines	KRAS	TP53	BRAF	PIK3CA
DLD1	G13D	S241F	wt	E545K; D549N
D-Mut1	G13D	S241F	wt	E545K; D549N
HCT116	G13D	wt	wt	H1047R
RKO	wt	wt	V600E	H1047R
SW48	wt	wt	wt	wt

wt, wild-type

5-Azacytidine, was purchased from Selleck Chemicals (cat: S1782). Stock solution was prepared in DMSO following the manufacturer´s instructions and was used in the assays at concentrations of 5 μM and 10 μM in DMSO solution. The control cells were treated with the same DMSO concentration to avoid unspecific effects.

3.2 Patient population

Three different patient's cohorts were used in this study. The first, a commercial cohort from TissueScan CRC cDNA array (Ori Gene, cat: HCRT101), the second cohort were

clinical samples obtained from patients of the Hospital Clínico de Valencia (HCV). Finally, the third is an *in silico* validation cohort, obtained through the public databases.

The TissueScan CRC cDNA array's cohort contained samples distributed as follows: 5 cDNA samples from histologically normal mucosa, 10 cDNA from patients with stage I, 13 cDNA from patients with stage II, 14 cDNA from patients with stage III and 6 cDNAs from patients of stage IV. The age of patients ranges from 36 to 92 years old, having an average age of 68.5 years with a deviation of 13.3 years. On the other hand, in relation to the sex of the patients, the sample is distributed in 21 men and 27 women.

A prospective cohort of 101 patients with CRC from HCV was used as a second and main study cohort (Table 3). This study has been approved by the HCV ethics committee, number 2017/229. The informed consents were signed by the patients before surgery. Of the HCV patient cohort, 100% samples were included in paraffin, with tumour tissue and healthy pairing. On the other hand, fresh frozen samples were available from 77 of these patients, all of them with localized disease.

A cohort of 622 patients from the The Cancer Genome Atlas (TCGA) database have been studied. On the other hand, 50 paired samples (CRC tissue and the healthy adjacent) of patients from the TCGA database have been compared. This last cohort of paired samples is composed of 8 stages I, 24 stages II, 9 stages III and 9 stages IV.

Table 3: Clinicopathological characteristics of the HCV patient cohort.

Clinicopathological characteristics of the cohort n (%)

Age average	73 (range 37-96)	
Gender	Male	61 (60.3%)
	Female	40 (39.6%)
T	T1	9 (8.9%)
	T2	17 (16.8%)
	T3	30 (29.7%)
	T4	26 (25.7%)
	nd	19 (18.8%)
N	N0	47 (45.5%)
	N1	22 (21.8%)
	N2	13 (12.9%)
	nd	19 (18.8%)
M	M0	82 (81.2%)
	M1	19 (18.8%)
Stage AJCC	I	12 (11.9%)
	II	35 (34.7%)
	III	35 (34.7%)
	IV	19 (18.8%)

nd: non-determinated

3.3 Genomic DNA extraction from cultured cells

Genomic DNA was extracted using the classical method described by Sambrook *et al*. (1989). Cells ($3x10^6$) were lysed in 500µl digestion buffer (10 mM Tris-HCl pH 8,0; 100 mM NaCl; 0,5% SDS (p/v); 25 mM EDTA pH 8,0; 100 µg/ml Proteinasa K) and incubated overnight a 50ºC under shaking. DNA was purified with 1 volume of phenol-chloroform-isoamyl alcohol (25:24:1) added to the lysed cell extract. After centrifuging for 10 min at 1,700g, the aqueous phase was transferred to a new tube. DNA was precipitated with 1/2 volume of 7,5 M ammonium acetate and 2 volumes of 100% ethanol, washed with 70% ethanol, dried and suspended in deionized sterile water.

Total DNA isolated was quantified in a Nano-Drop 2000 spectrophotometer (Thermo Fisher Scientific).

3.4 Total RNA extraction

For cell lines, the RNA extractions were carried out using the High Pure RNA Isolation Kit (Roche, Basilea, Suiza, cat: 11828665001) following the manufacturers´ protocol. The RNA extraction from samples embedded in paraffin was performed by RNeasy FFPE RNA extraction kit (Qiagen, cat: 73504). Paraffin samples were cut with a microtome, making 4 cuts per sample of a thickness of 20 μm each and the deparaffination solution (Qiagen, cat: 19093) was used to eliminate the paraffin from the tissue samples.

The checking for purity and quantification of the RNA were performed by NanoDrop 2000 (Thermo Fisher Scientific). As a consensus protocol of quality in the laboratory, the samples with absorbance ratios 260/280 lower than 1.7, indicating a contamination from aromatic compounds, were re-extracted and quantified. The same protocol has been followed in the case of samples with absorbance values in the ratio 260/230 lower than 1.5, corresponding with contamination with salts, carbohydrates or phenols, repeating the procedure described above.

3.5 Quantitative reverse transcriptase-PCR assays (RT-qPCR)

To carry out reverse transcription an initial amount of 500 ng of RNA was agreed upon. The reaction was carried by High-Capacity RNA-to-cDNA kit (Thermo Fisher Scientific, cat: 4387406) following the manufacturer´s instructions. The samples were then stored at 4°C, for an immediate use, or at -20 °C until required.

A real time PCR (qPCR) was performed using iTaq Universal SYBR Green Supermix (Bio-Rad) with a “QuantStudio 5 Real-Time PCR System” (Applied Biosystems). The relative value of expression was obtained using the 2^-ΔCt method where ΔCt was the Ct value of the sample minus the Ct value of the endogenous control. The amplification program used and the final melting curve analysis, to verify the specificity of the amplification product, is shown in figure 16. In 384-well plates, 9 μl of the reaction mixture of the qPCR and 1 μl of the cDNA sample were introduced, obtaining a final volume of 10 μl per reaction. The sequences of the primers used are shown in table 4.

Figure 16: Program for qPCR analysis of gene expression using QuantStudio 5 Real-Time PCR System and SybrG preparation software to SybrGreen assay.

Table 4: Primers sequence description used for expression studies.

Target	Forward oligonucleotide (5´- 3´)	Reverse oligonucleotide (5´- 3´)	Source
EPDR1 total	TGAAACCTGGATTGGCATCTATAC	TGTAGTTTATGGTAAAGGTTTCCTG	own
EPDR1 isoform1	GAGAGGAAGGCGCTGATCC	TGGCTTGGTCAATCTGAAACA	Riffo et al.
EPDR1 isoform2	GAGGAGGGTCTCTTGGGGAT	GCTGGGTGTTACTGAGTCCC	Riffo et al.
ZNF518B total	GGGCCTGAGGTTGTGAAACT	AAAACCGTGGCAAGTCCCAT	own
ZNF518B isoform1	GCTACAGGCAGGAATGTTACC	CGCAGTAGGTGCATGATCCC	Riffo et al.
ZNF518B isoform2	CTGCCGGTGTTACCTGGAAT	GCGCAGCTACTTCTTGGGT	Riffo et al.
ACTB	GTGCTATCCCTGTACGCCTC	GAGGGCATACCCCTCGTAGA	own
DDAH1	CAAAAGGACAAATCAACGAGGTG	TGTGCAGATTCACTAGACCCAA	Primer Bank
SNAI1	ACCACTATGCCGCGCTCTT	GGTCGTAGGGCTGCTGGAA	Primer Bank
CDH1	GTCAGTTCAGACTCCAGCCC	AAATTCACTCTGCCCAGGACG	Primer Bank
RBL2	GGAGGAAATTGGGACTCTCTCA	AGACGACTCAAGCTATGCGTA	Primer Bank
CREB5	TGGGACACATGATGGAGATGA	GGTGTGCATGAAGGTGGGAA	Primer Bank
SRPX	ATCAAGGTGAAGTATGGGGATGT	GTTTGACTGGCAGATCAGTAGG	Primer Bank
THBS1	GCCATCCGCACTAACTACATT	TCCGTTGTGATAGCATAGGGG	Primer Bank
NRP1	ACGTGGAAGTCTTCGATGGAG	CACCATGTGTTTCGTAGTCAGA	Primer Bank
BCAM	GAGGTGCGCTTGTCTGTACC	GCATATAATGGTCGTGGGTTCC	Primer Bank
TOMM34	TGCATCAAAGATTGCACTTCAGC	GCAGCACAGTCTTATAGTCAACA	Primer Bank
PTBP2	TGCAGTTGGCGTGAAGAGAG	ATGCTGCTCATATTAGAGTTCGG	Primer Bank
S100A14	GAGACGCTGACCCCTTCTG	CTTGGCCGCTTCTCCAATCA	Primer Bank

Primer Bank: Wang et al. 2009[191]

Riffo et al, 2016[158]

For the evaluation of the expression of microRNA 193a-5p (Thermo Fisher Scientific, cat: A25576) the methodology used varies slightly. As a housekeeping gene RNU43 (Thermo Fisher Scientific, cat: 001095) was used. A reverse transcription was previously performed using the TaqMan MicroRNA Reverse Transcription Kit (Applied

biosystem, cat: 4366596) following manufacturer´s instructions. Quantitative PCR was performed using “TaqMan Universal Master Mix II not UNG” (Applied biosystem, cat: 4440040). The amplification program required for the qPCR is shown in figure 17.

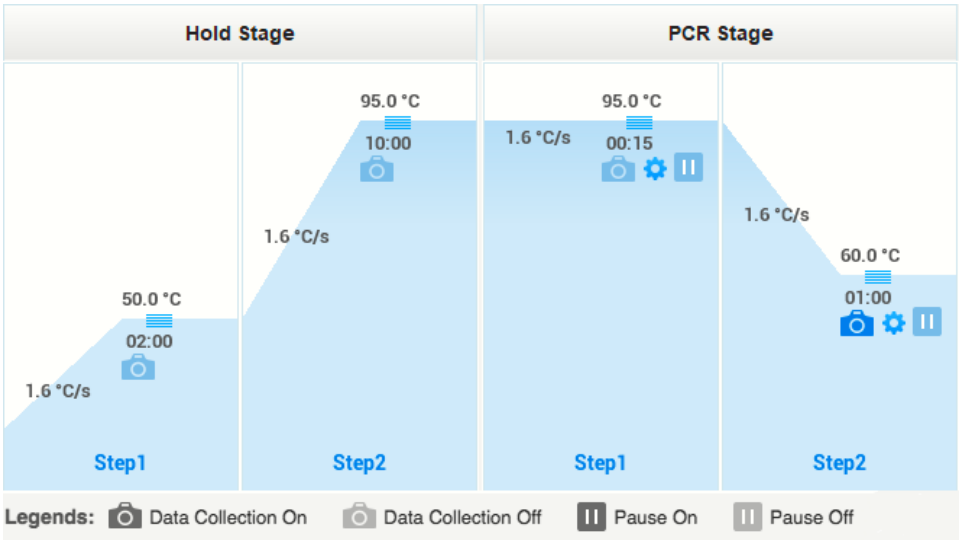

Figure 17: Program for qPCR analysis of gene expression using QuantStudio 5 Real-Time PCR System and TaqMan probe for the assay.

3.6 Protein extraction from cultured cells

To obtain total protein extracts sub-confluents cell cultures were washed with PBS and lysed with RIPA buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% Triton X-100, 0.1% SDS, 0.5% deoxycholic acid sodium salt) supplemented with 2 µl/ml protease inhibitor cocktail (Sigma Aldrich). Samples were incubated on ice 15 min, sonicated with 1 pulse of 10 seconds at an amplitude of 37% on an Ultrasonic homogenizer (Fisherbrand Model 120 Sonic Dismembrator, Thermo Fisher Scientific) and centrifuged at 13,000×*g* for 30 min at 4ºC. Total protein was determined by the BCA protein assay kit (Thermo Fisher Scientific; cat: 23225). On the other hand, nuclear protein extractions were carried out using the nuclear extraction kit (Active Motif, cat:

40410) following the manufacturer's recommendations. For the nuclear protein quantification Bradford (Ref) method is used.

3.7 Immunoblot analysis

Similar amounts of proteins were solved in loading buffer (Bio-Rad). Proteins were resolved by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) varying the percentage of acrylamide, according to the needs of the experiment (12% to 6%). Then, a transfer was made to nitrocellulose membranes (Bio-Rad) using the Mini-protean II system (Bio-Rad). The membranes were blocked for 1h at room temperature in 5% (w/v) BSA-containing TTBS (20 mM Tris-Hcl pH 7.6; 137 mM NaCl, 0,1% Tween 20) on a shaking platform and incubated overnight at 4ºC with the primary antibodies. The antibodies used are described in the Table 5. Specific bands were recognized using peroxidase-conjugated secondary antibody (Dako, cat: P0260). Immunoblots were visualized using the ECL Western Blotting detection kit reagent (GE Healthcare) and the ImageQuant LAAS 400 (Healthcare Bio-Sciencies) system. Protein levels were normalized by GADPH expression for total extracts or HDAC1 for nuclear extracts. For a semi-quantitative determination of protein in the western blots, four grey values relative to the loading control were measured with ImageJ and averaged.

Table 5: Antibodies and blotting conditions.

Antibody	Dilution	Supplier	Reference
N-CADHERIN	1/500	Cell Signaling Technology	13116
SNAIL	1/500	Cell Signaling Technology	3879
ZNF518B	1/100	Sigma Aldrich	HPA031216
GADPH	1/2000	Abcam	ab-8245
HDAC1	1/300	Santa Cruz	sc-8410

3.8 Immunostaining of fixed samples

To detect EPDR1 and ZNF518B proteins expression in cultured cell lines an immunocytochemistry analysis was performed. 50,000 cells in 500 μl of culture medium were seeded per well on a plate culture cell slide (SPL Life Sciences, cat: 30108). After the appropriate washes, the cells were fixed with 4% paraformaldehyde and permeabilized with 2% Triton X100. Normal horse serum 2.5% (Vector, cat: S2012) was used as a blocking solution during 1 h. Then, cells were incubated with 1/200 dilution of either EPDR1 (Abcam, cat: ab197932) or ZNF518B (Sigma, cat: HPA031216) antibodies. An anti-rabbit secondary antibody (Dako Agilent, cat: P0448) was used at 1/100 dilution and it was revealed with DAB (Dako Agilent, cat: 3467). The reaction was stopped for the subsequent haematoxylin-eosin staining.

Likewise, for detection of the EPDR1 protein expression in patient samples, it was performed an immunohistochemistry analysis on both tumour and non-tumour paired tissues. The analysis was performed on 5 μm paraffin sections, after which the sample was hydrated by means of several washes with decreasing alcohol concentrations, and the sample was autoclaved for retrieval at low pH (Dako Agilent,

cat: 10128667). After washing, peroxidase was blocked (Dako Agilent, cat: S2023), then the primary EPDR1 antibody was added (1/200) in the diluting solution (Dako Agilent, cat: S0809). Finally, after incubation with the primary antibody and washings, the anti-rabbit secondary antibody is added at a 1/100 dilution. It was developed with DAB and the reaction was stopped for the subsequent haematoxylin-eosin staining.

To assemble the sample, it is dehydrated with several washes with increasing alcohol concentrations and mounted with Entellan (Merck KGaA, cat: 1.07961.0500).

3.9 *EPDR1* and *ZNF518B* gene expression knockdown by transfection of synthetic small interfering RNAs

Knockdown expression of *EDPR*1, was carried out with two specific siRNA (si1) (Qiagen, cat: SI04235721) and siRNA (s2) (Qiagen, cat: SI03238053). In the case of *ZNF518B* it was used an equimolar mixture of two siRNAs (Qiagen, cat: SI04131015 and SI04284805) or a transfection with each of them separately. In all cases, a transfection with siRNA control (Qiagen, cat: 1027280) was carried out to assess specific effects of *EPDR1 or ZNF518B* siRNAs.

Cells were seeded in 6-well plates one-day prior transfection to reach a confluence of 50-60% in a medium without antibiotics. Cells were transfected using a final concentration of 20 nM of siRNAs and Lipofectamine RNAi Max (Invitrogen, cat: 1875252) as a transfection reagent following the manufacturer´s instructions. The efficiency on the silencing were analysed by RT-qPCR after 48 and 96 hours of transfection.

All the experiments for the evaluation of the functionality of EPDR1 and ZNF518B derived from the modification of their gene expression have been repeated and replicated experimentally in order to obtain significant differences.

3.10 Overexpression of *EPDR1* and *ZNF518B* by transfection with cDNA expression plasmids

Overexpression was carried out by transient transfection using plasmids as expressing vectors in cell lines that do not express *ZNF518B* and *EPDR1*, such as SW48 and RKO. The plasmid required for the overexpression of *EPDR1,* pCMV3-*EPDR1*-GFPSpark (Sino Biological, HG13665-ACG) and for the overexpression of ZNF518B, pCMV3-*ZNF518B*, have the constitutive CMV (immediate-early cytomegalovirus) promoter. As a control plasmid for both cases an empty CAGGS plasmid was used (Oxford genetics, cat: OG504R1), plasmid donated by Dra. García-Murria, Department of Biochemistry of the University of Valencia.

Plasmids used for transfection were amplified in *E. coli* and purify with QIAprep Spin Miniprep Kit (Qiagen, cat: 27104) according to the manufacturer's protocol. The purified plasmids were quantified with a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific) and stored at -80ºC until used.

One day prior transfection, $3x10^5$ SW48 and RKO cells were plated in 6 wells plate, and let them grow in a medium without antibiotics until an 80% confluence. Then, a mixture of 4 μg of plasmid and 10 μl of Lipofectamine 2000 Reagent (Invitrogen, cat: 11668-027) was added in Opti-MEM Medium (Thermo Fisher Waltham, cat:

31985047). After 6 h, the global medium was changed for maintenance up to 24 h with standard medium.

3.11 Cell growth assay (MTT)

Cells were seeded in 96-well plates at a rate of 3,000 cells per well in quadruplicate after 24 h of transfecting with siRNAs. For the evaluation of cell proliferation, 10 µl of a 5 mg/ml solution of MTT (3- (4,5-dimethylthiazol-2-yl-2,5-diphenyltetrazolium bromide) were added into each well containing 100 µl of culture medium at several times after seeding. After incubation for 4 h at 37 °C, 80 µl of the content was poured off and 200 µl of dimethyl sulfoxide (DMSO) was added. After that, the plate was read in the spectrophotometer at a wavelength of 560 nm.

3.12 Transwell migration and invasion assays

The migration and invasion assays were carried out using transwell plates of 12 wells (Hampton, cat: 08-771-21), with pores of 8 µm in size. For the invasion assays, an amount of 50 µl of matrigel (Sigma Aldrich, cat: 3445-010-01) was added to the transwell inserts, conferring an extracellular matrix mimicking substance and imitating an *in vivo* invasive process. For both migration and invasion, 10^5 cells were seeded in upper compartment of the inserts after 24 h of transfection with siRNAs or plasmids. Foetal bovine serum was used as an attractive substance for the cells to pass through the pores. The migration test was stopped 24 h after the introduction of the cells in the inserts, whereas the invasion test was stopped after 48 h. The cells located at the

lower side of the porous membrane of the well, where fixed with 70% methanol and stained with 0.2% violet crystal. The counting of the stained cells was carried out with the help of an inverted microscope (DMi8, Leica). Six different regions of the porous membrane of the well were counted, so that a mean and a deviation of the count could be obtained.

3.13 Wound-healing assay

The analysis of the cells migration capacity, dependent on the expression of the *EPDR1* and *ZNF518B* genes, was also carried out by a wound healing test through the *Radius 24 Cell Migration Assay kit*" (Cell Biolabs, cat: CBA-125). For this test 24 h post-transfection 8,000 cells were seeded in each well allowing them to adhere to the plate for up to 24 h. Then, they were seeded in each well allowing them to adhere to the plate for up to 24 h. After that, and following the indications of the manufacturer´s, the gel layer was eliminated from the central zone of the well, so that a free central space without cells was obtained. Migration of the cells through the empty zone was monitored by photographing the critical zones at 24, 48 and 72 hours.

3.14 Colony formation assays

The cells were seeded in 6-well plates 24 h post-transfection with siRNAs or plasmids. In order to observe proliferation from single cells, only 150 cells per well were used. After 2-week culture, the medium was removed; cells were washed with PBS, fixed with 70% methanol, and stained with 0.2% violet crystal. Finally, around 50 colonies

were counted at different points of the well with the help of an inverted microscope (DMi8, Leica, Wetzlar, Germany).

3.15 Cell adhesion assays to type I collagen coated plates

To study the capacity of adhesion to type I collagen fibres as a result of the expression changes of the *EDPR1* and *ZNF518B* genes, 60 mm Petri dishes coated with collagen type I were used (Corning, cat: 62405-423). 30.000 cells were seeded 24 h post-transfection in their corresponding medium, and after 30 min, the medium was eliminated and the plates were washed with PBS, so that only cells that have been able to adhere to the collagen I coated plate could be counted. Six different regions of the plate were counted.

3.16 Flow cytometry analysis of cell cycle distribution

To evaluate cell cycle changes related to EPDR1 or ZNF518B knockdown in DLD1 and HCT116 cell lines,-flow cytometry assays were done. After 72 h of silencing, cells were trypsinized and incubated with a hypotonic propidium iodide solution a 4ºC for 12 h. The samples were subsequently analysed by cytometry (Gallio, Beckman Coulter) in the Cytometry Service from the Central Unit of Research in Medicine, UV-INCLIVA. The distribution of the cells in the different phases of the cycle were calculated using the FloJo software (TOMY Digital Biology). Each experiment was repeated at least three times.

3.17 DNA methylation analysis

The analysis of DNA methylation was carried out through AGENA's MassARRAY platform technology (AGENA Biotechnology) in the Epigenetics and Genotyping Service from the Central Unit of Research in Medicine, UV-INCLIVA. To carry out the methylation analysis, amplification by PCR is required in which the T7 RNA polymerase promoter is labelled from a DNA sample modified with bisulphite. This will generate a chain of monocatenary RNA derived from the modification with bisulphite, then proceeded to the excision of the chain by RNase A by cutting specifically in UTP or CTP. The primers used to determinate the methylation are shown in table 6. Primers 1 and 2 were required for the analysis of the intron upstream of exon 3, covering an amplicon of 347 and 226 bp respectively. Primers 3 amplify a fragment of 338 bp at the 5' non-translatable end of exon 3. Primers 4 were used for the study of methylation of the translatable region of exon 3 and its downstream intron, generating a 326 bp amplicon.

Table 6: Primers sequence description used for analysis of DNA methylation of EPDR1.

Primers	Forward	Reverse	Size (bp)
1	TAGGTTGGATGGAGTGTAGTGGTAT	ACTCCCATTCCCAATAAAAAATCTA	347
2	TTTAGATTTTTTATTGGGAATGGGA	TCAACAAATAATCACACAAAAAAAA	226
3	TTTTTTTTGTGTGATTATTTGTTGA	AAACCAAAATACTTC CTATCCCTTC	338
4	GTTTTTTTGGGTTTGGGTTTGT	AACCTATACCTCTTACAA AAAATCAAC	326

After this process, a set of products of different sizes and mass were analysed by MALDI-TOF-MS mass spectrometry (TripleTOF 5600+ System, SCIEX, Alcobendas, Madrid). In this way, the abundance of each specific fragment analysed by mass spectrometry provides a quantitative evaluation of the methylation of each CpG.

3.18 Chromatin structure and epigenetic modification analysis

3.18.1 Isolation of mononucleosomes

To isolate mononucleosomes from CRC cell lines, the protocol described in Riffo *et al*. 2014[192] were followed. Briefly, cells were fixed with 1% formaldehyde for 5 min, centrifuged at 500 xg for 5 min, washed with PBS, resuspended in 6 volumes of cell lysis buffer (100 mM NaCl, 3 mM $MgCl_2$, 30 mM sucrose,10 mM EDTA, 0.5% Nonidet P-40, 10 mM Tris-HCl, pH 7), supplemented with 2 μl/ml protease inhibitor cocktail (Sigma Aldrich, San Luis, MI, USA), and incubated on ice during 15 min. After incubation, the suspension was centrifuged at 1000 xg for 5 min, and the nuclear pellet was resuspended in wash buffer (15 mM Tris-HCl pH 7,5; 15 mM NaCl; 3 mM $MgCl_2$; 60 mM KCl; 20% (v/v) glycerol). Nuclei were digested with 25 units of micrococcal nuclease (MNase) (Roche Applied Science) per A260 unit in wash buffer supplemented with 3 mM $CaCl_2$. Digestion was carried out during 15 min at 37 ºC and stopped by adding EDTA to 10 mM. Then, the digestion was treated with 5 μg/ml RNase and 200 μg/ml of proteinase K for 1 h at 60ºC and mononucleosomal DNA was purified using the GeneJET PCR purification kit (Fermentas). The intact nuclear was digested with micrococcal nuclease to yield a preparation of mononucleosomes.

3.18.2 Nucleosome occupancy

To estimate the position of nucleosomes, isolated nuclei from cross-linked cells were incubated with MNase as describe above to obtain mononucleosomes. After purification of the mononucleosomal DNA, qPCR analysis was performed using primers (Table 7) that amplified short tiled fragments of about 100 bp covering the

region comprised between from -1,000 to +300 relative to the transcription start site (TSS) of *ZNF518B*. The results for each primer pair in the qPCR reaction were normalized using the data obtained with sonicated genomic DNA of an average size of 300pb. The normalized data were plotted against the position of the amplicon centre. The experimental results were compared with the output of the sequence-based prediction of positioning carried out using the NuPoP software tool (https://www.bioconductor.org/packages/release/bioc/html/NuPoP.html). The entire experiments were repeated at least three times.

Table 7: Primers used for nucleosome occupancy determination and Nuc-ChIP.

Amplicon*	Forward	Reverse	Size (pb)
-1012	CAACTCCGCTTCTCCGTGT	CATCACCTTCCTGTTGGCGG	110
-915	GTGATGGAAACCAGCCTTGC	CAGCGTCACTGGGAAGTACA	96
-845	ACTGTACTTCCCAGTGACGC	GCACAATCAGTGACTCCCCA	87
-775	GAGTCACTGATTGTGCGGGA	TACGGTCCGGAAGGAAGACA	89
-715	TATGTCTTCCTTCCGGACCGT	TTCGGTGCTATTTTGGGCGA	71
-651	GTTGGCTCGCCCAAAATAGC	AGGGTCAGAGGGGACAATGA	110
-580	CATTGTCCCCTCTGACCCTC	GGGAAGGTCGTATGTGAGGC	70
-521	CTCACATACGACCTTCCCCG	TCTTAGCTGCTGGACCTTGG	84
-453	TTTCTCGCCAAGGTCCAGC	TCGTCTTCTTGGAGAGCTTCAG	107
-385	ACTGAAGCTCTCCAAGAAGACG	GTGGGTGAGTGAAAGCGAGA	74
-324	TCTCGCTTTCACTCACCCAC	CCTGCCTTGTACCGGGTC	88
-249	GACCCGGTACAAGGCAGG	AGCGAGGTCCACATTTAGCC	98
-165	GGCTAAATGTGGACCTCGCT	TTCTAAGCTGTCCAGAGCGG	109
-108	TCTGAGGCTCCCCGCTCT	CTGAGGTCTCTAAGGCTGCG	67
-80	CCGCTCTGGACAGCTTAGAA	GCGTGCGCAATGTGAAGC	102
-27	TCAGGCGCCGGCCTCGCTGGAG	CGCCGCGCCCGCTGTAGGTC	104
-16	CGCCGCTTCACATTGCG	GCCCGCTGTAGGTCCCTA	70
+6	GCCGCCGCTTCACATTGC	GGCGAAGGGGCGTCTACA	117
+79	CGGTGTAGACGCCCCTTC	CCTCCCGCTCAGCTACTTAC	71
+111	GACCTGGCACCCGAATC	CGCTTCGGTCACGTACAC	86
+120	TGCCGGGTAAGTAGCTGAG	GCGCTTCGGTCACGTACAC	65
+156	GAGGGCCCCCGGATCCGC	GTCAGGCCCCCGCCCCGC	118
+199	TGACCGAAGCGCCAGC	GACGCCAGCCCTCAGC	119
+272	CGTGCCCGCGCTGAGGGC	GCTGGTGGGCGTGGGGCA	78

Description of the sequence of the 24 amplifications required to cover the nearby areas and the site of initiation of transcription (TSS). * the negative numbers represent the distance downstream of the TSS (upper part of the table), the positive ones being the upstream areas (lower part of the table).

3.18.3 Analysis of histone epigenetic modifications by mononucleosomal immunoprecipitation (Nuc-ChIP)

Epigenetic modifications of histones were studied at nucleosomal level by mononucleosomal immunoprecipitation approach (Nuc-ChIP) as previously described in Riffo *et al*. 2014[192] and Riffo, *et al*. 2018[146]. The following antibodies were used: anti-H3K9me3 (Abcam, Cambridge, UK, ab-8898); anti-H3K4me3 (Abcam, ab-8580); anti H3K9ac (Abcam, ab-4441); anti-H3K27ac (Abcam, ab-4729). Antibodies were conjugated with Dynabeads Protein G (Invitrogen) adding the optimal quantity of each antibody to immunoprecipitate 3,5 units of chromatin per sample.

To determine the quantification of the H3K9ac and H3K27ac tags, the amplicons -521, -324 and -108 were required, which cover respectively the N-3, N-2 and N-1 nucleosomes. On the other hand, H3K4me3 and H3K9me3 were analysed on the +75 amplicon, which covers the initial part of nucleosome N+1 in the cell lines DLD1 and D-Mut1, and the final part of the nucleosome that occupies the TSS in the SW48 cell line.

3.19 Gene expression microarray analysis

To study the effects of *ZNF518B* knockdown on the transcriptomic profile of CRC cell lines, the Clariom^{TM} S Assay containing more than 20,000 well-annotated genes (Affimetrix Thermofisher) was used. Cell RNA was obtained as described above after 72 h of siRNA transfection and it was processed in the Transcription Unit of Central Unit of Medical Research (UCIM) of the University of Valencia.

The experiment was carried out in triplicate to ensure the reliability and reproducibility of the results in the two cell lines HCT116 and DLD1. The analysis of the data was carried out through the Transcriptome Analysis Console bioinformatics program (TAC) (Thermo Fisher Scientific). To obtain a list of genes with significant changes (p-value < 0.05) between control and silencing conditions a filtering algorithm was applied. The algorithm uses a combined criterion, being the fold-change of both conditions calculated through the log2 of the replicates average.

Several gene expression changes from the array were validated by RT-qPCR. A total of 10 primers were used as shown in table 4. Functional annotation of differentially expressed genes in the cell line models was performed using the Pathway Studio V10.

3.20 Statistical analyses

The statistical analysis was carried out in consequence to the number of variables to be compared. The distribution of the data was analysed by the normality tests of *Shapiro-Wilk* and *Kolmogorov-Smirnov*. On the other hand, the equality of variances was analysed by the *Levene* test. All the statistical analyses were carried out wih the software *RStudio* (version 1.0.136) and *GraphPad Prism6*. By consensus, all p-values less than 0.05 are considered significant.

In the case of statistical analysis where the comparison between two variables was made, a statistical test was required, such as *Student's t* test. On the other hand, when the analysis involved the comparison of more than two variables, a *Kruskal-Wallis* or *ANOVA* test was used according to the data normality.

The differential expression of *EDPR1* between the healthy tissue and the tumour was determined by two methods. First, it was carried out through the difference of expression, subtracting the expressions between tumour tissue and peripheral tissue of the same patient, and second, as a quotient of the result between the division of the expression values of tissues (fold-change). The expression of *EPDR1* in the two tissues was compared by a *t-test* of paired samples. In the part of the studies only with fresh sample, *t-test* of non-paired samples was used for its evaluation. Likewise, for the study of the relationship between the difference of expression of *EPDR1* or fold-change, and the different predictive variables such as variances T, N, M and stages AJCC, a multiple regression analysis was used.

The relationship between the survival time of the patients and the covariates, which takes into account the expression in tumour and peripheral tissue and the covariates T, N, M and AJCC stages, was studied through a *Cox regression* analysis.

RESULTS

Results

4.1 Study of *EPDR1* gene as a putative biomarker in colorectal cancer progression.

4.1.1 EPDR1 expression in patients with colorectal cancer

4.1.1.1 The EPDR1 gene and its isoform 2 are overexpressed in tumour samples of colorectal cancer with respect to healthy samples

The expression of *EPDR1* in CRC was first analysed in a commercial cDNA array. To do this, a qPCR analysis of the expression of total *EPDR1* and of its isoform 2 was performed in an OriGene TissueScan array of cDNA.

Expression of total *EPDR1* is detected in normal mucosa at a very low level, but is clearly expressed in almost all the tumour samples. Only in 5 out of the 43 tumour samples the fold change of *EPDR1* expression relative to the normal mucosa is less than 5. As shown in Figure 18, the upregulation of the whole gene is found in all the CRC stages. The analysis of the gene expression using the usual ANOVA approach and multiple comparisons, revealed a significant difference ($p = 0.041$) between normal and tumour tissues. The behaviour of isoform 2 was somewhat different. First of all, as previously reported in CRC cell lines[158], its expression level is low, and it is undetectable in normal tissues. Moreover, independently of the stage, the plots showed that in some CRC patients isoform 2 is not detected or expressed at a negligible level, while there are other patients showing a higher expression (Figure 18). Stage II patients show a large dispersion of expression values, both in whole *EPDR1* and in its isoform 2.

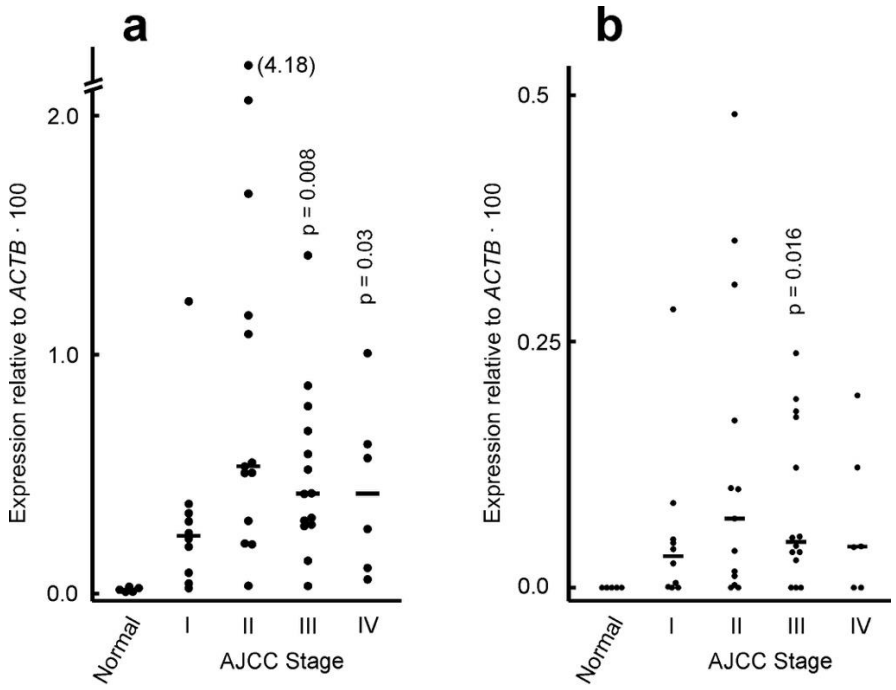

Figure 18: Expression of *EPDR1* in a cDNA array of normal colonic mucosa and tumours measured by qPCR. The tumour samples were classified according to the AJCC stage. (**a**) Expression of total *EPDR1*. (**b**) Expression of *EPDR1* isoform 2. The p-values are given when a significant difference with normal mucosa is found. The position of median values is shown.

4.1.1.2 *EPDR1* gene is overexpressed in colorectal cancer patients from a prospective cohort from the Hospital Clínico de Valencia

The expression of *EPDR1* was then studied in a prospective cohort of 101 CRC patients included in the study in the last 3 years. The characteristics of the cohort have been given under Materials and Methods. The expression of *EPDR1* was determined by RT-qPCR after extraction of the RNA from the paraffin-embedded samples. Out of the 101 samples, 4 were discarded since the values indicative of RNA quality were not optimal for the subsequent analysis. Figure 19 shows that there is a significant increase (p-value <0.05) in the expression of *EPDR1* in the tumour tissue when compared to the healthy peritumoural tissues of the 97 paired samples.

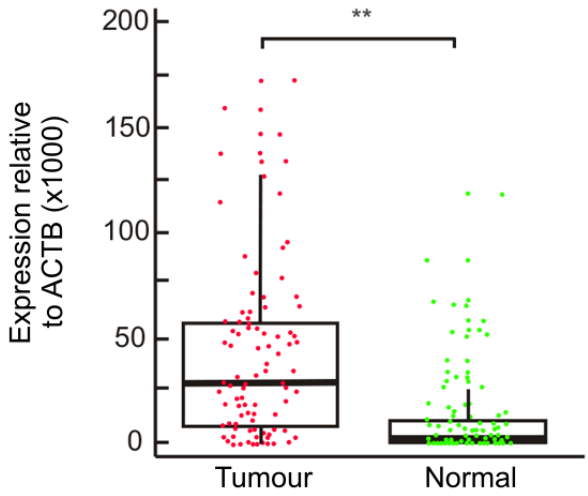

Figure 19: Box plots of total *EPDR1* expression in our cohort of 101 CRC patients of the HCV. *EPDR1* expression was measured by RT-qPCR, using *ACTB* as reference, in tumour and non-tumour adjacent tissues. Outliers with expression values higher than 200 (4 from tumours and 3 from non tumour samples) are not included in the graph, but they are shown in figure 28. **, $p < 0.01$.

The distribution of the gene expression in the samples is clearly visualized in a density graph (Figure 20) where the red line marks the profile of the distribution of expression in the tumour samples. The profile shows a wide distribution of different values, and the average of the tumour values (red arrow) is clearly high with respect to healthy tissues (line and blue arrow). On the other hand, the line of distribution of densities corresponding to the expression of healthy tissues is significantly displaced to the left of the graph, which indicates a very high number of samples with very low or even no expression of *EPDR1* in these tissues. It should be highlighted that Figure 20 shows the existence of small peaks in the blue line, indicating the presence of a few healthy samples whose expression of *EPDR1* is over the average of the expression in healthy subjects, and even above the average of the expression in

tumour samples. These samples of healthy tissue correspond in their entirety to patients in stage IV of the development of the disease.

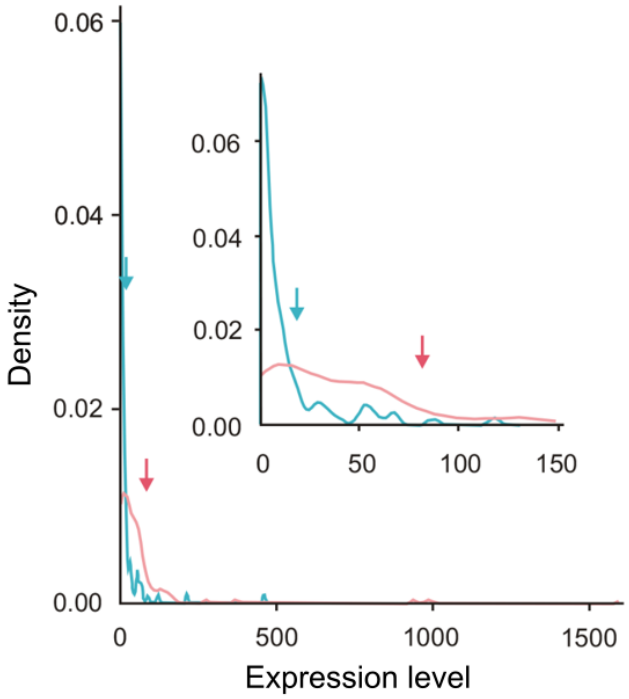

Figure 20: Density plot of *EPDR1* expression in samples from CRC patients. The red line represents density *versus* level of expression of *EPDR1* in the tumour tissues. The blue line represents density *versus* level of expression of *EPDR1* in peripheral tissues. The arrows represent the averages of gene expression in each type of tissue. The inset shows a magnification of abscissa scale of the plot to appreciate the differences between the two tissue types for expression values lower than 150.

The expression of *EPDR1* in the tumour tissues and in their corresponding paired healthy samples is represented in Figure 21. In this way it is possible to analyse and compare data from paired samples of the same patient. With few exceptions, *EPDR1* expression is significantly higher in tumour tissue than in tissue peripheral to the tumour in both localized (Figure 21A) and metastatic disease (Figure 21B) (p-value <0.05). Note that the ordinate scales of the graphs of Figure 21 are different, as the

expression of *EDPR1* in metastatic patients is much higher than that of patients with localized disease.

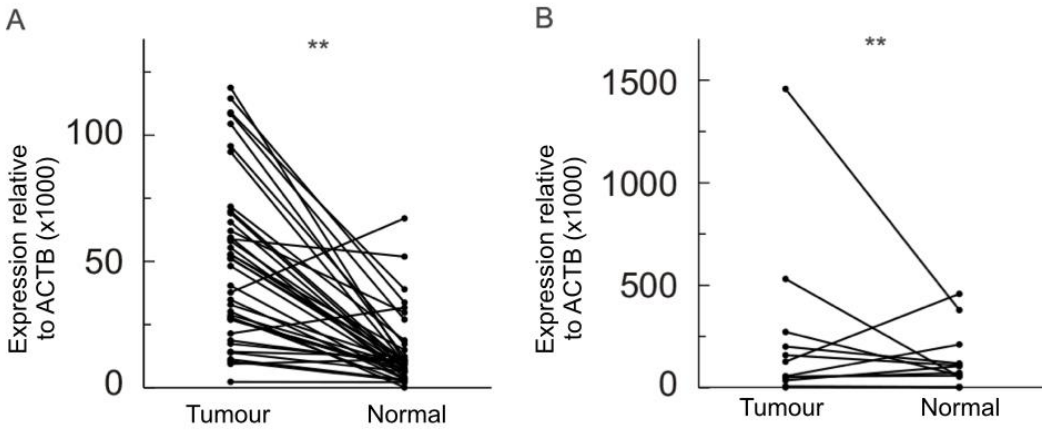

Figure 21: Total *EPDR1* gene expression measured by RT-qPCR in tumour samples compared to their paired normal tissue. A) Samples of patients with localized CRC. B) Samples from patients with metastatic CRC. **, $p < 0.01$.

4.1.1.3 Colorectal cancer staging variables as predictors of EPDR1 expression

To determine whether the staging variables are predictors of the *EPDR1* expression, different statistical studies, described in the Material and Methods section, were performed. *EPDR1* expression relative to *ACTB* was grouped with respect to the TNM parameters, as shown in Figure 22. The expression of the *EDPR1* gene in all the AJCC, T and N stages show significant differences when compared to normal tissues by the Student's t test (Figure 22 A-C). The significance, as revealed by the p-values, is greater in the late AJCC stages (III and IV). However, the p-values of the differences between *EPDR1* expression in cancer and healthy tissues with respect to the parameters T (Figure 22B) and N (Figure 22C) are maintained.

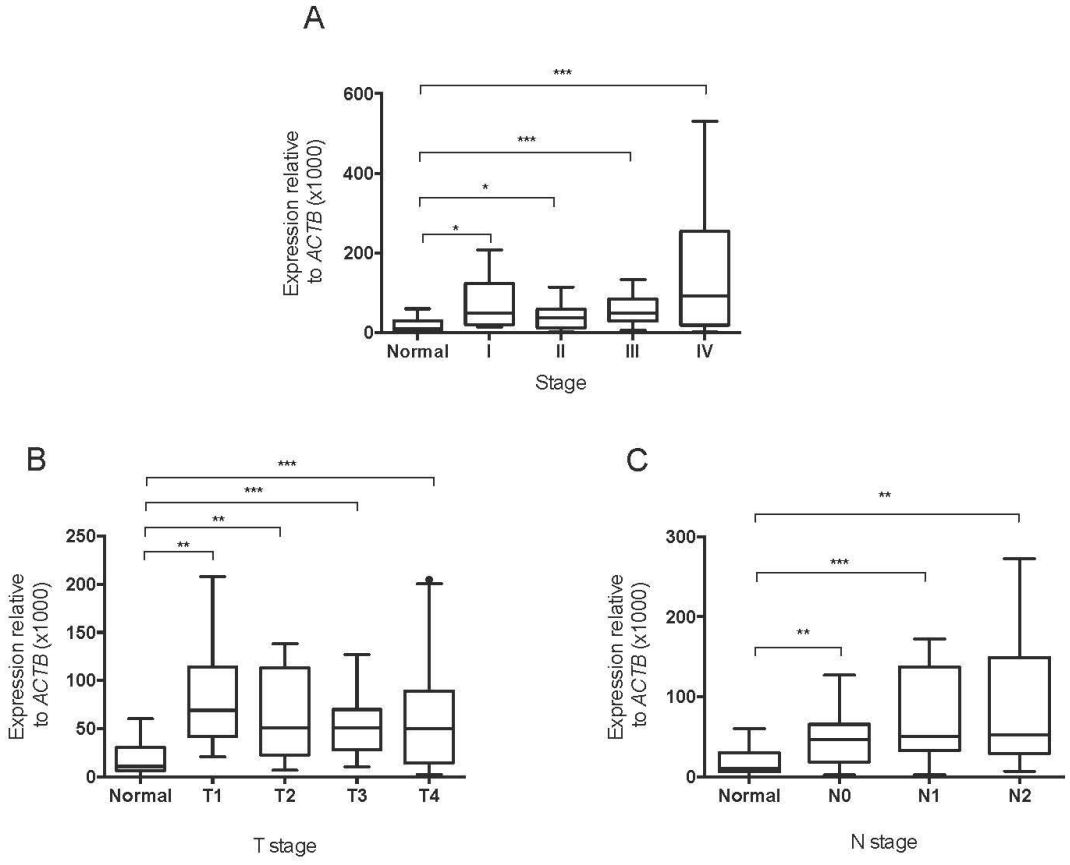

Figure 22: Total *EPDR1* gene expression measured by RT-qPCR in patients with CRC. A) Expression of *EPDR1* in relation to the AJCC stages of the disease. B) Expression of *EPDR1* regarding the distribution of patients according to the parameter T. C) Expression of *EPDR1* with respect to the distribution of patients according to parameter N. **, $p < 0.01$. ***, $p < 0.001$

The possible relationship between the expression of *EPDR1* in the tumour tissue of patients and other clinicopathological parameters was also tested. The relation between the expression of *EPDR1* and the perineural invasion (Figure 23A), vascular invasion (Figure 23B), tumour differentiation (Figure 23C) or the presence/absence of incipient tumour budding (Figure 23D) were also included in the study.

In none of these cases clear significant differences were observed with respect to the expression of *EPDR1* (p-values > 0.05).

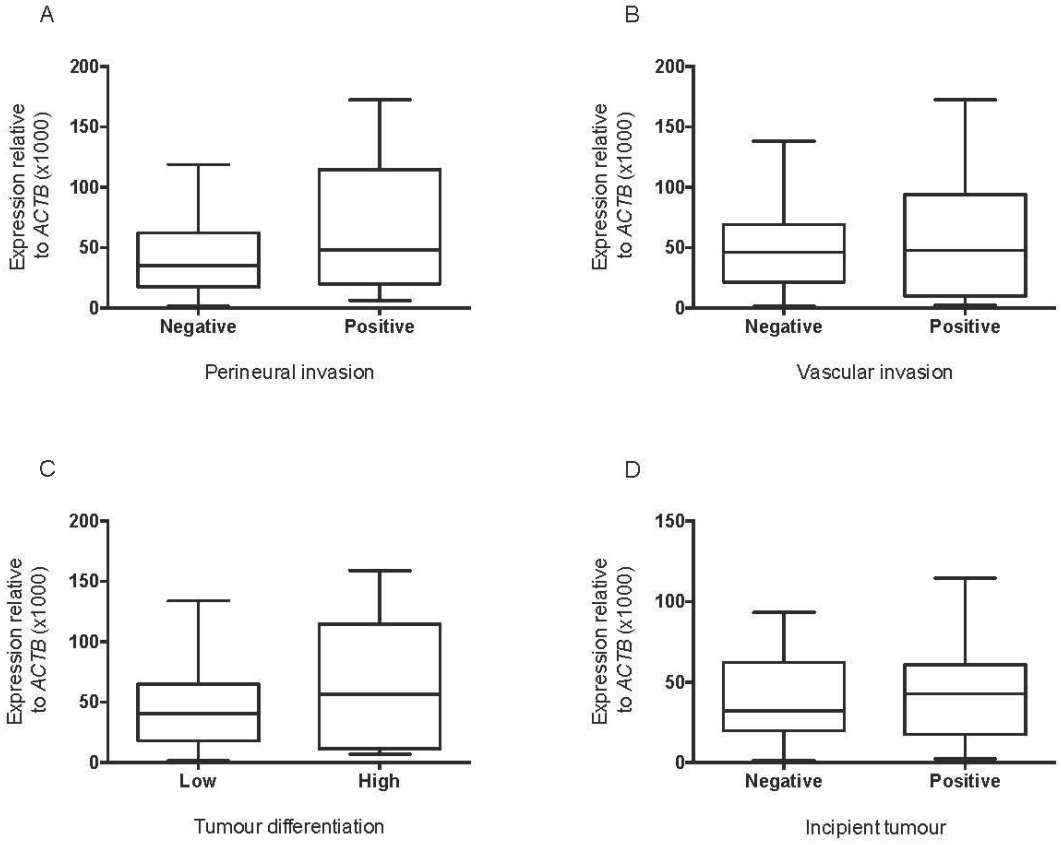

Figure 23: Expression of the *EDPR1* gene relative to *ACTB* according to different clinicopathological parameters of the patients. A) Expression of the *EDPR1* gene based on the distribution of patients relative to the perineural invasion of the tumour. B) Expression of *EPDR1* in relation to vascular invasion of the tumour. C) Expression of *EPDR1* based on the degree of tumour differentiation in the patients. D) Assessment of the relationship between the expression of *EPDR1* and tumour budding.

A more exhaustive study, based on a multiple regression model, was then carried out to correlate the expression of *EPDR1* with the TNM and AJCC parameters. In this model all the parameters as a whole were taken into account as predictive values of gene expression. Statistical calculations were made taking into account the final expression value of *EDPR1* obtained in two ways. The first of them was the value

of the ratio resulting from the division of *EDPR1* expression in the tumour by that of the non-tumour, further referred to as fold change. The second way was simply the difference of the *EPDR1* expression between the tumour and the corresponding peripheral part of each patient. As mentioned, in both cases, the statistical analysis was based on a multiple regression model.

When the analysis included the TNM and the AJCC stages, a significant coefficient in the ratio tumour/adjacent expression was observed with T3 and T4 stages (p-value = 0.0021 and p-value = 0.0009, respectively), with M1 stage (p-value = 0.0004) and with AJCC stage II (p-value = 0.0005). No significant coefficients were observed between the ratios and N stages or with AJCC stage III. The p-value of the test for null coefficients corresponding to the variables T2, T3 and T4 is 0.00038. When the categorical predictors corresponding to N and M stages were removed from the model, a significant correlation with AJCC stages III and IV ($p = 0.0013$ and $p = 4.8 \cdot 10^{-7}$) was also observed. All these results, summarized in Table 8, manifest that a strong relation exist between *EPDR1* expression in CRC tissues and the invasiveness of tumour cells.

When the analysis was done starting from the difference of *EPDR1* expression between the tumour and the peripheral region, there were no great differences relative to the previous analysis, but it has to be highlighted that the coefficients obtained with the N parameter, both in N1 and N2 patients were significant (p-value <0.05 in both cases). Nevertheless, the parameter N does not affect the significance in stage III (table 8).

Table 8: Summary of the results obtained by the statistical analysis of the expression of the *EPDR1* gene in relation to the TNM and AJCC parameters by means of a multiple regression model.

Parameters		Expression ratio	Model p-value	Global p-value	Expression difference	Model p-value	Global p-value
T	**T1**	32,003		0,0003	36,457		$8,0x10^{-5}$
	T2	247,121	0,9969		171,332	0,9292	
	T3	65,057	0,0021		45,833	0,0014	
	T4	41,005	0,0009		26,697	0,0005	
N	**N0**	59,194		0,2258	44,437		0.0356
	N1	135,319	0,1162		95,265	0,0148	
	N2	117,561	0,0869		80,234	0,0113	
M	**M0**	31,200		0,0004	33,084		0.0066
	M1	358,940	0,0004		206,376	0,0066	
AJCC	**I**	31,707		0,0020	40,776		0.0007
	II	28,262	0,0005		28,839	0,0003	
	III	33,087	0,2062		31,172	0,5734	
	IV	358,940	NA*		206,376	NA*	

* Value not available for the required statistical model.

4.1.1.4 The expression of *EPDR1* as a predictor in patient survival

The survival probability of the study cohort from HCV was estimated by means of a Kaplan-Meier analysis. It was calculated the dependence on the survival of patients with the expression of *EPDR1* by means of a Cox regression analysis. The result of the expression of *EPDR1* in tumours in relation to the survival of patients is close to significance (p-value =0.085), but it should be noted a clearly significant dependence (p-value =0.0057) between the expression of *EPDR1* in tissues adjacent to the tumour and their survival.

When a Cox regression including the AJCC stage was carried out, the survival time was also found significantly dependent on the *EPDR1* expression in tissues adjacent to the tumour (p-value =0.0344) but not on the expression in the tumour itself (p-value =0.1161). Similar results were obtained if the Cox regression included the T stage. In this case significant dependence on the expression in non-tumour adjacent tissue was found (p-value =0.0127) but no significant dependence on the expression in tumour was found (p-value= 0.2188). In other words, the expression of *EPDR1* in the environment of the tumour tissues seems to be a better predictor of survival time than the expression in the tumour itself.

4.1.1.5 *In silico* analysis corroborates the overexpression of *EPDR1* in two independent cohorts of patients.

To compare the results obtained with the HCV and the *OriGene TissueScan array* patients' cohorts, an *in silico* analysis was performed using data from TCGA. The *EPDR1* expression is included in the TCGA database in a cohort of 622 patients with CRC. Figure 24 shows the descriptive Kaplan-Meier curve of the cohort, which shows the probability of survival over time of the 622 patients selected for the *in silico* study.

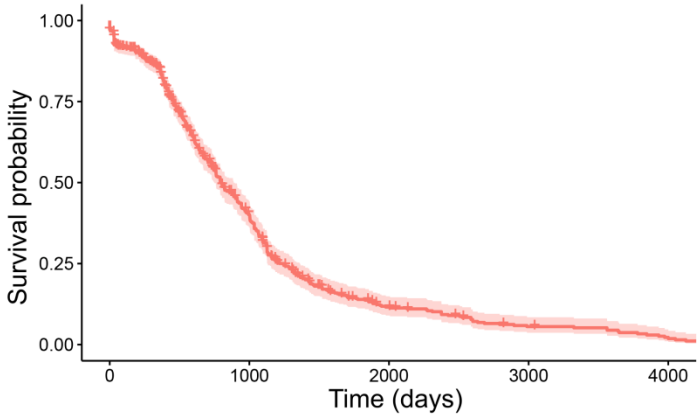

Figure 24: Kaplan-Meier curve descriptive of the survival probability of the TCGA cohort of 622 patients with CRC.

It was possible to verify the relationship of the expression of *EDPR1* with the different stages in which, the 622 patients of the *in silico* cohort were classified. A Cox regression test showed a clear and significant differences in the expression of EPDR1 with survival in those patients classified in stage II (p-value =0.00177), but differences were not observed in stages III and IV.

The comparison between the expression of *EPDR1* in tumour samples and the expression of the gene in healthy samples from TCGA was in agreement with the above reported results. When the 622 patients' data were analysed, the expression of *EPDR1* in tumour samples was 2.08-fold higher than in healthy samples, with a clear significant difference (p-value <0.0001) analysed using a Welch two-sample t-test. On the other hand, the comparison carried out only with paired samples from TCGA (with a sample size of 50) also agrees with present results, since the expression in tumour was 1.7 times higher than in the adjacent paired tissue (Table 9). The fold-change of gene expression within each stage was significantly different in stages II, III and IV (p-value <0.05 in all cases).

Table 9: *In silico* analysis of the expression of *EPDR1* in paired samples of CRC recovered from TCGA database.

Stage	Sample size	Expression ratio	p-value
All	50	1.705	3.631×10^{-5}
Stage I	8	1.421	0,328
Stage II	24	1.614	0.0068
Stage III	9	2.645	0.0038
Stage IV	9	1.743	0.0176

A T-test of paired samples was used to perform the analysis. In the table is shown the sample size of patients used and the relationship between tumour tissue expression and the paired normal tissue.

A second *in silico* study was performed using data from RNA-seq paired sample's cohort reported by Kim *et al*. 2014[193].These authors generated a series of data from 18 CRC metastatic patients in which they included samples from normal mucosa, primary tumours and metastasis. This publicly available database (https://www.ncbi.nlm.nih.gov/bioproject/218851) was used to recover the *EPDR1* expression values using the same bioinformatic procedure described for our group in Riffo *et al*., 2016[158]. The gene expression in primary tumours was 2.15-fold increased over than in normal mucosa. Interestingly, a higher fold change (2.518) was observed when metastases are compared with normal mucosa (Table 10). Therefore, all the results acquired *in silico,* both from TCGA and from the Kim *et al.* data, agreed with those obtained from our EPDR1 expression data from both the local HCV and OriGene TissueScan array patients' cohort studied.

Table 10: An *in silico* analysis of *EPDR1* expression in paired samples of CRC from RNA-seq data published by Kim *et al*. 2014 [193].

Paired samples	Sample size	Expression ratio	p-value
Primary tumour vs. normal mucosa	18	2.150	7.9×10^{-5}
Metastasis vs normal mucosa	18	2.518	4.1×10^{-6}

A T-test of paired samples was used to perform the analysis. In the table is shown the sample size of patients used and the relationship between tumour tissue expression and the paired normal tissue.

4.1.1.6 Ratio of EPDR1 Isoforms as a predictor of malignancy and survival in colorectal cancer fresh samples

All the above results refer to the whole gene expression, which is mainly accounted for by isoform 1[158]. Next, the expression of the minor isoform 2 was studied in the local patients' cohort. To do this, fresh tumour tissues should be used, because the quality of RNA recovered from paraffin-embedded samples does not allow determining that scarcely abundant isoform. Fresh tissue was only available from 77 samples of HCV's cohort. As expected, isoform 1 is significantly more expressed than isoform 2 in this cohort of patients (Figure 25, p-value <0.001). Interestingly, some changes in the ratio isoform 2 / isoform 1 (ratio i2/i1) were observed in some patients. The isoform 2 is scarcely expressed with respect to the isoform 1, but a small increase of the isoform 2 with respect to the 1 could change the prognosis of the patient. Note that the differences in the isoform 2 expression while subtle, can be clearly enough to observed.

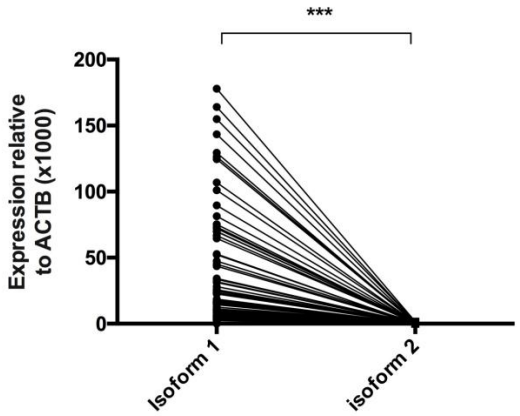

Figure 25: Expression of isoforms 1 and 2 of the *EPDR1* gene in 77 fresh samples of the HCV's cohort. The difference of expression was statistically significant with a p-value< 0,001, analysed by T-test of paired sample.

The role of *EPDR1* isoforms are unknown so it is difficult to determine whether there is a functional effect for this alternative splicing alteration. In the last few years, alterations in alternative splicing are emerging as relevant drivers of cancer[194]. In this sense, we decided to examine whether a correlation exists in our 77 HCV patient's cohort between the ratio of i2/i1 and the different clinico-pathological parameters.

Figure 26 A shows that the ratio isoform 2/isoform 1 significantly increases as the parameter T increases. In other words, the increase in isoform 2 with respect to 1 is significantly related to an increase in the infiltrating capacity and tumour size. Contrarily, there are no significant differences in the ratio with respect to lymph node infiltration (Figure 26 B). As to the isoform ratio changes with the AJCC stage, it can be clearly seen that the more advanced the stage of the disease the higher the expression ratio (Figure 26 C). Another interesting result of these experiments is that an increased i2/i1 ratio is related to a decrease of the probability of disease-free survival, since a significant increase in the ratio is observed in relapsing patients (Figure 26 D). The results given in Figure 26 E are particularly remarkable. The isoforms ratio has been compared with the presence of *KRAS* mutations in the patients and it is plainly seen that a positive relation exists between the mutational state and the ratio. The vascular invasion of the tumour does not correlate with the isoform ratio (Figure 26 F), although in the peritoneal invasion (Figure 26 G) there is a significant relationship with isoforms ratio. Interestingly, in this instance, it can be observed that the higher the ratio the lower the perineural invasion.

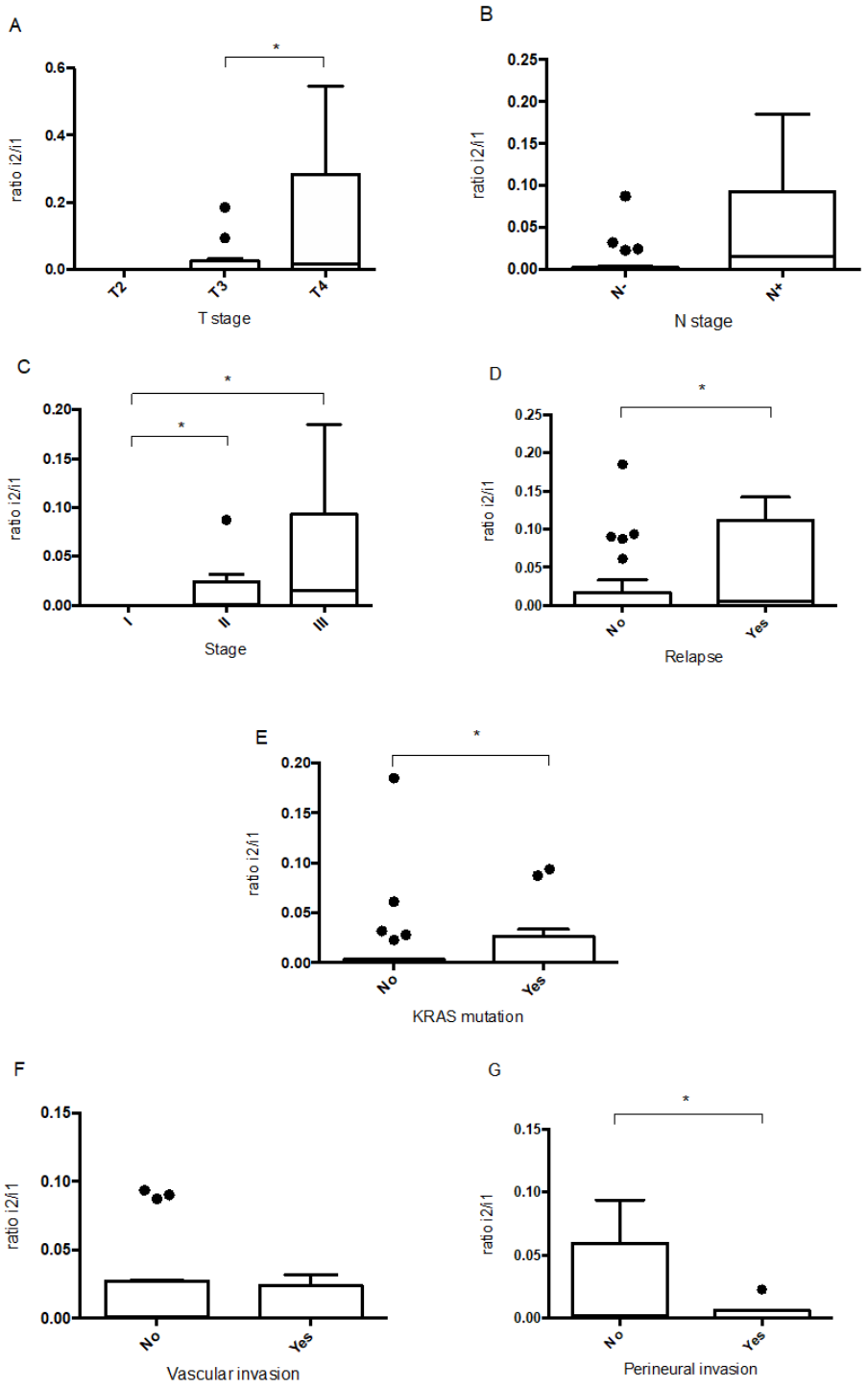

Figure 26: Representation of the ratio of expression between isoform 2 and isoform 1 of *EPDR1* with respect to different clinicopathological parameters. The analysis was carried out with the fresh tissue samples of the cohort. Ratio i2/i1 of *EPDR1* regarding the distribution of patients according to: A) the parameter T; B) the parameter N; C) the stages I, II and III. D) absence or presence of relapse in the disease; E) absence or presence of mutation in *KRAS*; F) vascular invasion; G) Perineural invasion. In A and C panels a Kruskal-Wallis test was performed, in rest of the panels a T-test statics was used. *, $p < 0.05$.

The interesting results given in Figure 26 D suggested an additional analysis, namely the determination of the disease-free survival probability of patients according to their ratio between *EPDR1* isoforms. To do this, the patients were first divided into two groups, further referred to as “low ratio” and “high ratio”. The “low ratio” group, defined as those samples with isoform ratio lower than the median value of the ratios, and the “high ratio”, which have a ratio higher than the median. Then a Cox regression was done and the results were plotted as a Kaplan-Meier graph (Figure 27). It is clear that the higher the ratio i2/i1, the lower the disease-free survival of the patients.

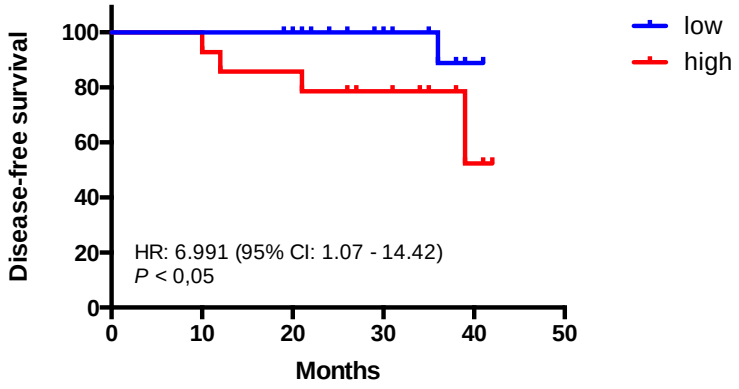

Figure 27: Kaplan-Meier representation of the relapse probability of the HCV patient cohort based on the differences between the ratio of isoforms 2 and 1 of *EPDR1*. Cutting data for differentiation between "low" and "high" is determined based on the median of the overall data. Cox regression analysis showed a p-value < 0.05 and a HR of 6.991 with a 95% of CI between 1.07 and 14.42

4.1.1.7 The protein EDPR1 is overexpressed in colorectal cancer samples

Preliminary immunohistochemical studies of paired samples from the local patients’ cohort showed that in tumour sample the staining with the EPDR1 antibody was higher than in adjacent tissue (Figure 28). At a global level, the differences are clear, in the image corresponding to the tumour tissue an intense cytoplasmic / membranous

staining is observed in the epithelial cells. In contrast, weak cytoplasmic staining of crypt epithelial cells is observed in normal tissue, and weak stromal staining is also observed. As can be observed in the amplifications of the micrograph of figure 28, the staining mainly concentrated in the cytoplasm, being fainter and even null in the nucleus, effects due to non-specificities of the antibody.

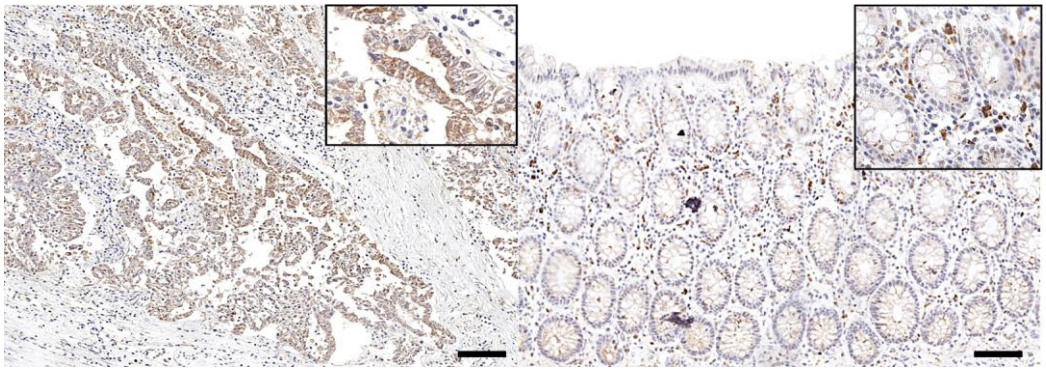

Figure 28: Immunohistochemistry of EPDR1 in patient sample with CRC (on the left) and its corresponding healthy counterpart (on the right). the black scale line represents 150 µm.

4.1.2 Functional analysis of the EPDR1 gene in colorectal cancer cell line models

4.1.2.1 Knockdown of *EPDR1* gene using small RNA interference (siRNAs) molecules in two colorectal cancer cell lines

The above results show that the expression of whole *EPDR1* and/or of its isoforms, is in some way related to development and progression of human CRC *in vivo*. To explore the mechanisms involved in this *EPDR1*-related development of CRC, it is important to know the implications of the gene in cell proliferation and other phenotypic properties in human CRC cell lines. For this purpose, one of the strategies employed in this study was the silencing of *EPDR1* gene by siRNAs. In a previous work, our group demonstrated that two CRC cell lines, DLD1 and HCTT16 express this gene,

so they were selected for this analysis. After transfection the efficiency of silencing was checked by RT-qPCR (figure 29). The results observed in figure 37 were obtained with one specific commercial siRNA, named siEPDR1 (s1).

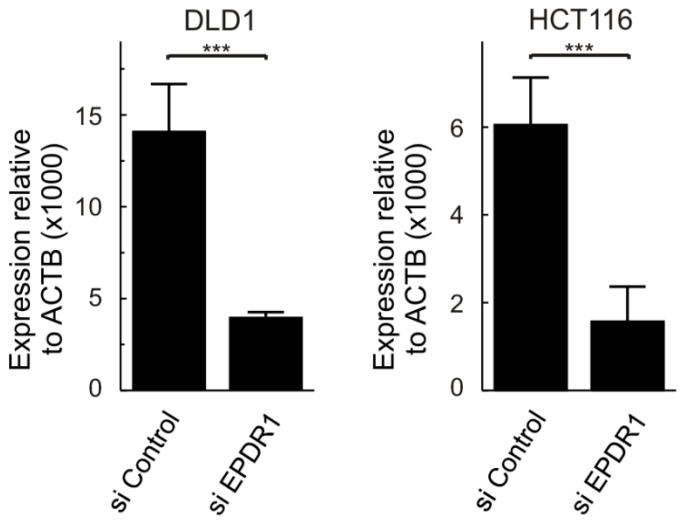

Figure 29: Expression of the *EDPR1* gene after silencing with the *EPDR1* specific siRNA (si1) in CRC cell lines DLD1 and HCT116. The results obtained are represented against the control, using a scramble siRNA for it. The statistical analysis required for the comparison of the two variables was a Student t-test. ***, P < 0.001.

4.1.2.2 EPDR1 promotes cell proliferation in colorectal cancer cell lines

The consequences of *EPDR1* knockdown on cell growth were evaluated by MTT, colony formation assays and flow cytometry. The MTT assay shows that *EPDR1* has a significantly clear role (p-value <0.001) in cell proliferation (Figure 30). When the gene is silenced, both DLD1 and HCT116 are unable to proliferate, remaining with a constant population over time.

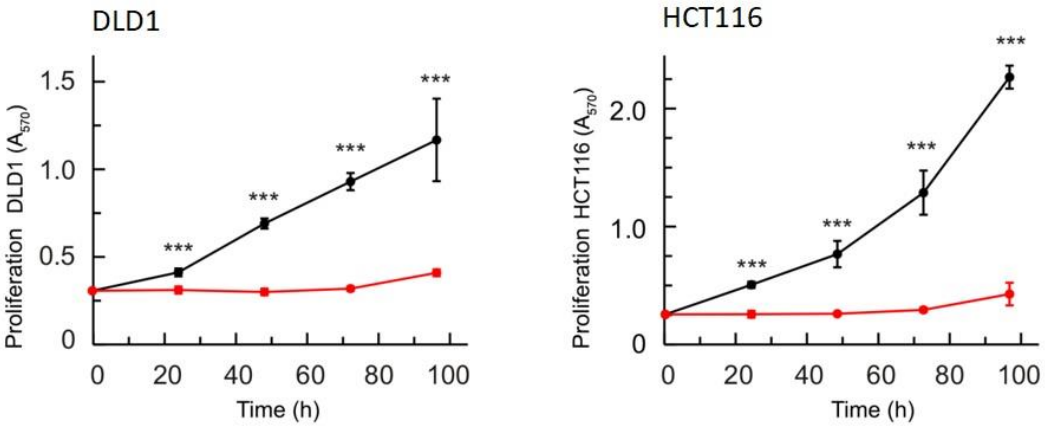

Figure 30: Effects of *EPDR1* silencing on cell proliferation measured by MTT test in two *EPDR1*-expressing cell lines, DLD1 and HCT116. The graph shows the results of two independent experiments, run in triplicate. The black line corresponds to the control cells transfected with scrambled siRNA, while the red corresponds to *EDPR1* knockdown cells transfected with si1. The results were compared by Student's t-test. ***, $p < 0.001$.

These results, as well as those of the colony formation test (Figure 31), allow concluding that *EPDR1* significantly influences the ability of CRC cell lines to proliferate and to not be affected by anoikis.

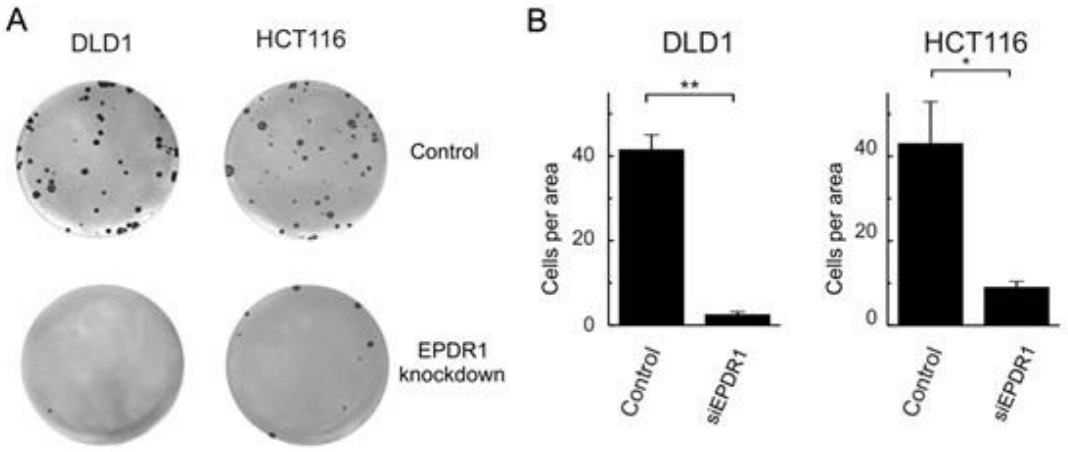

Figure 31: Effects of *EPDR1* silencing on the colony formation test in DLD1 and HCT116. A) Representative figures of plates showing the colonies that can be observed at the end point. B) Averaged cell number quantification of 3 fields in 3 plates each. The results were compared by Student's t-test. *, $p < 0.05$. **, $p < 0.01$.

To further explore the mode of action of *EPDR1* on the proliferation of CRC cells, the influence of silencing the gene on the distribution of cells into the different cell cycle phases was next studied. It should be noted that silencing *EPDR1* generates an accumulation of cells in the G0 / G1 phase, decreasing the ability of the cells to enter S phase. This effect can be seen in Figure 32, in both, DLD1 and HCT116, at 24 hours after reseeding silenced cells. It is worth emphasizing the dramatic increase in apoptosis, more pronounced in DLD1 than in HCT116, at 48 hours after reseeding. The percentage of cells in G2 / M phase does not vary after silencing of *EPDR1* at the different times in which the experiment was performed in any of the cell lines.

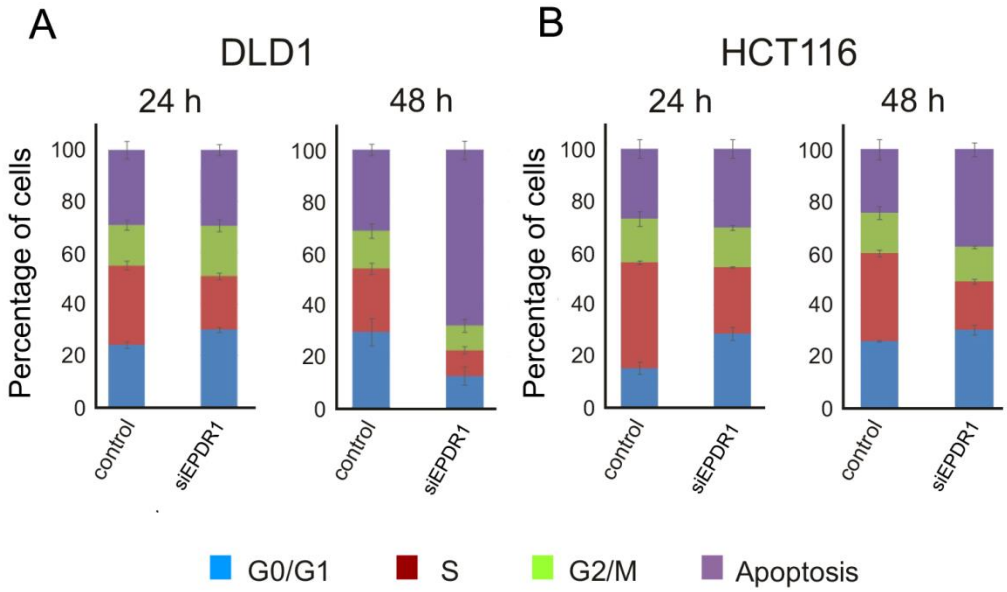

Figure 32: Flow-cytometry cell cycle analysis of EPDR1 silenced cells. (A) DLD1 cell line and (B) HCT116 cell line, transfected with scrambled siRNA (control) and with si1 (EPDR1 knockdown) at 24h and 48h after reseeding silenced cells.

4.1.2.3 *EPDR1* promotes cell migration, invasion and adhesion to collagen type I fibres

The effects of *EPDR1* knockdown on cell invasion, migration and adhesion to collagen type I was next analysed in DLD1 and HCT116 cell lines. These studies provide information on the possible implications of the *EDPR1* gene in the EMT, a crucial step in the development of cancer cells.

Migration of cells has been studied in two ways. The first one was the transwell assays, whose results indicate a significant reduction (p-value <0.001) in the migration capacity through the pores of the inserts of the cell lines DLD1 and HCT116 when the *EPDR1* gene is silenced (Figure 33).

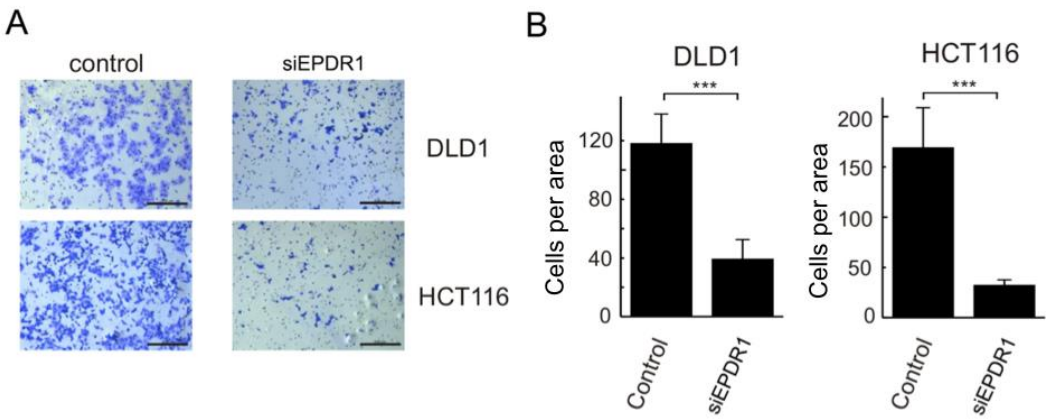

Figure 33: Transwell migration assay of cells transfected with scramble siRNA (control) and with si1 (*EPDR1* knockdown) in DLD1 and HCT116 cell lines. A) Photographs of representative plates after 24 h of migration are shown. The size bars correspond to 200 µm. B) Averaged quantification of the results of the migrated cells from 3 transwell plates. The results were compared by Student's t-test. ***, $p < 0.001$.

The second form of migration analysis was the wound healing method. It can be seen in Figure 34, that after 48 hours, in both cell lines, a clear effect of *EPDR1* silencing was observed. While under the control conditions roughly 95% of the gap was closed by migration and proliferation of cells, after *EPDR1* silencing, the situation was more

or less similar to that at 24 hours. At 72 hours, in the controls the gap had been completely closed, but after silencing *EPDR1* the gap closure was less than 50%. Both experiments show how *EPDR1* has an important role in the proliferation and migration capacity for the CRC cell lines, DLD1 and HCT116.

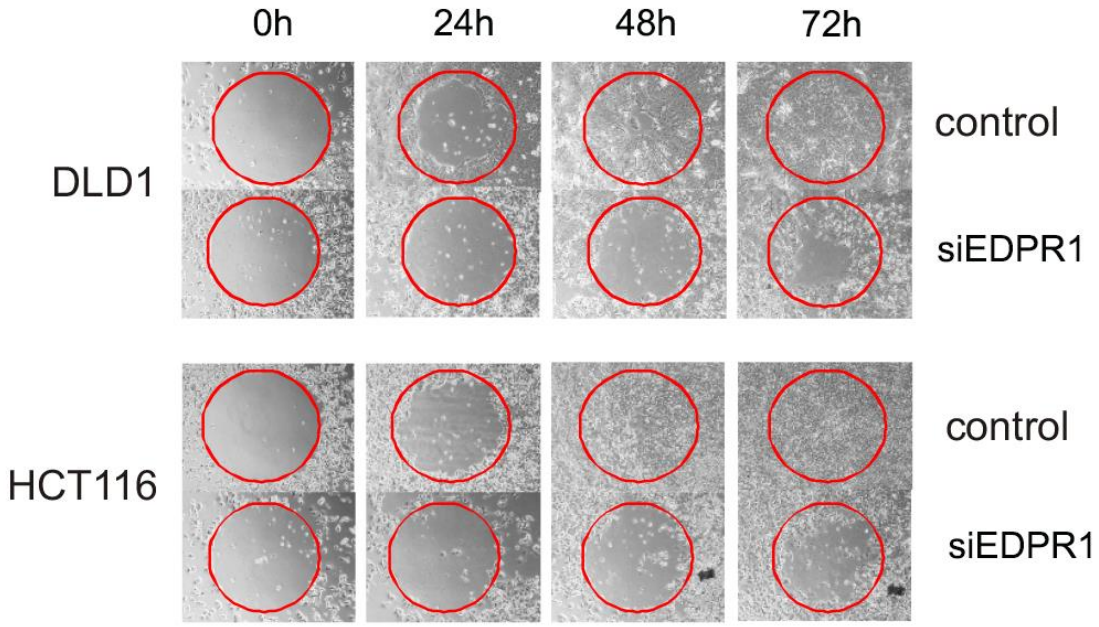

Figure 34: Wound healing test after silencing *EPDR1* in DLD1 and HCT116 cell lines. Cells were treated either with scrambled siRNA (control) or with si1 (*EPDR1* knockdown). The time of culture after removing the central gel layer (shown as a red circle) is given at the top.

To assess the involvement of *EPDR1* on the invasion capacity of DLD1 and HCT116, transwell experiments in which Matrigel was added on the porous membrane of the transwell inserts were carried out as indicated under Material and Methods. The counting of the cells capable of crossing the transwell membrane with the Matrigel layer (Figure 35) indicates that *EPDR1* silenced cells do not possess the same invasive ability than those normally expressing *EPDR1* with a clear significant difference (p-value <0.001).

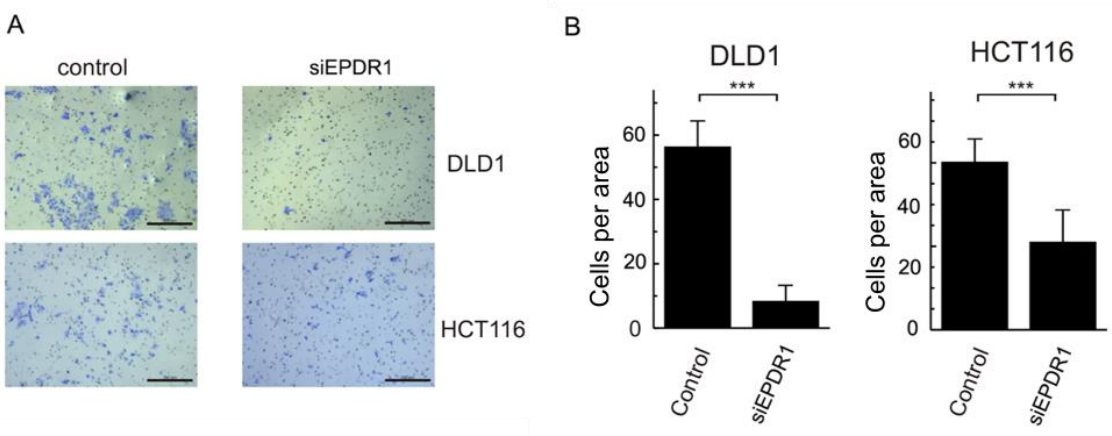

Figure 35: Effects of knocking down *EPDR1* expression on cell invasion capacity. A) Transwell migration assay through Matrigel layer of cells treated with scrambled siRNA (control) and with si1 (*EPDR1* knockdown). A photograph of representative plates is shown. The size bars correspond to 200 µm. B) Averaged quantification of the results of 3 transwell plates. The results were compared by Student's t-test. ***, $p < 0.001$.

Ependymins are known to adhere to collagen fibres[161]. To check whether this property is also shared by EPDR1 was next studied. As shown in Figure 36 after silencing the *EPDR1* gene, the adhesion capacity of the cell lines DLD1 and HCT116 is affected in a significant way (p-value <0.001). This means that *EPDR1* has a function in the adhesion to type I collagen fibres in those cell lines.

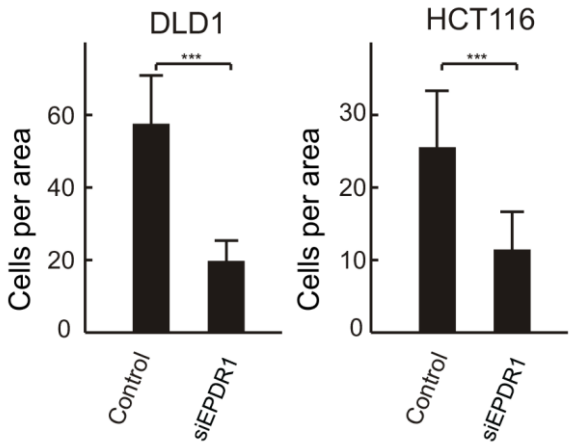

Figure 36: Adhesion test to type I collagen-coated plates. The effects of knocking down the gene *EPDR1* on the adhesion of cells was quantified and the plots show the average of three cell counts. The results were compared by Student's t-test. ***, $p < 0.001$.

4.1.2.4 Functional analyses with another siRNA-mediated knockdown of *EPDR1* (si2) exclude off-targets effects

To confirm the results described and to discard off-targets effects of the siRNA (si1) used above, *in vitro* experiments were carried out with another siRNA of *EPDR1* (si2).

Figure 37 shows that the two siRNA used referred as si1 and si2 were similarly efficients in knocking down the transcription of the gene.

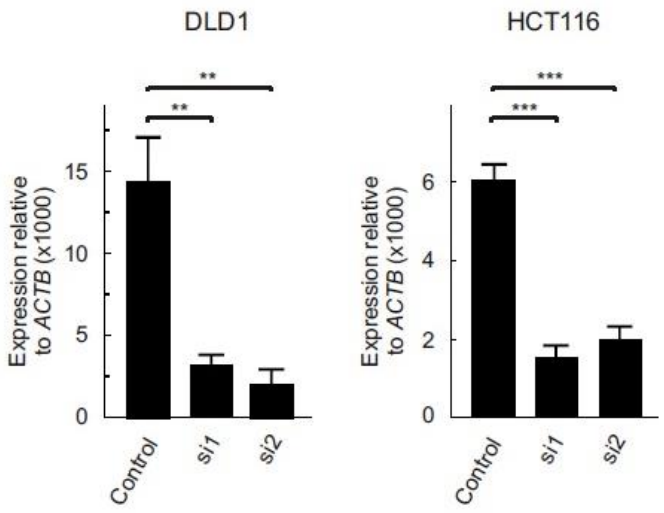

Figure 37: Comparison of the efficiency of *EPDR1* silencing in DLD1 and HCT116 cells with two siRNAs, si1 and si2. Expression level of *EPDR1* relative to that of the β-actin gene was measured by RT-qPCR after transfection of cells either, with scrambled siRNA (control) or with specific siRNAs (si1 and si2), as described under Material and Methods. The results were compared by Student's t-test. **, $p < 0.005$; ***, $p < 0.001$.

The effects observed in figure 38 with the si2 of *EPDR1* corroborate the results obtained with si1 in relation to proliferation. The consequences of this knockdown were evaluated by MTT, colony formation assays and flow cytometry. After *EPDR1* silencing the proliferation of cells was almost negligible in MTT and colony formation assays compared with cells treated with scrambled siRNA. Flow cytometry assay confirm the effect of accumulation of cells in G0/G1 phase after *EPDR1* silencing like the results obtained with si1.

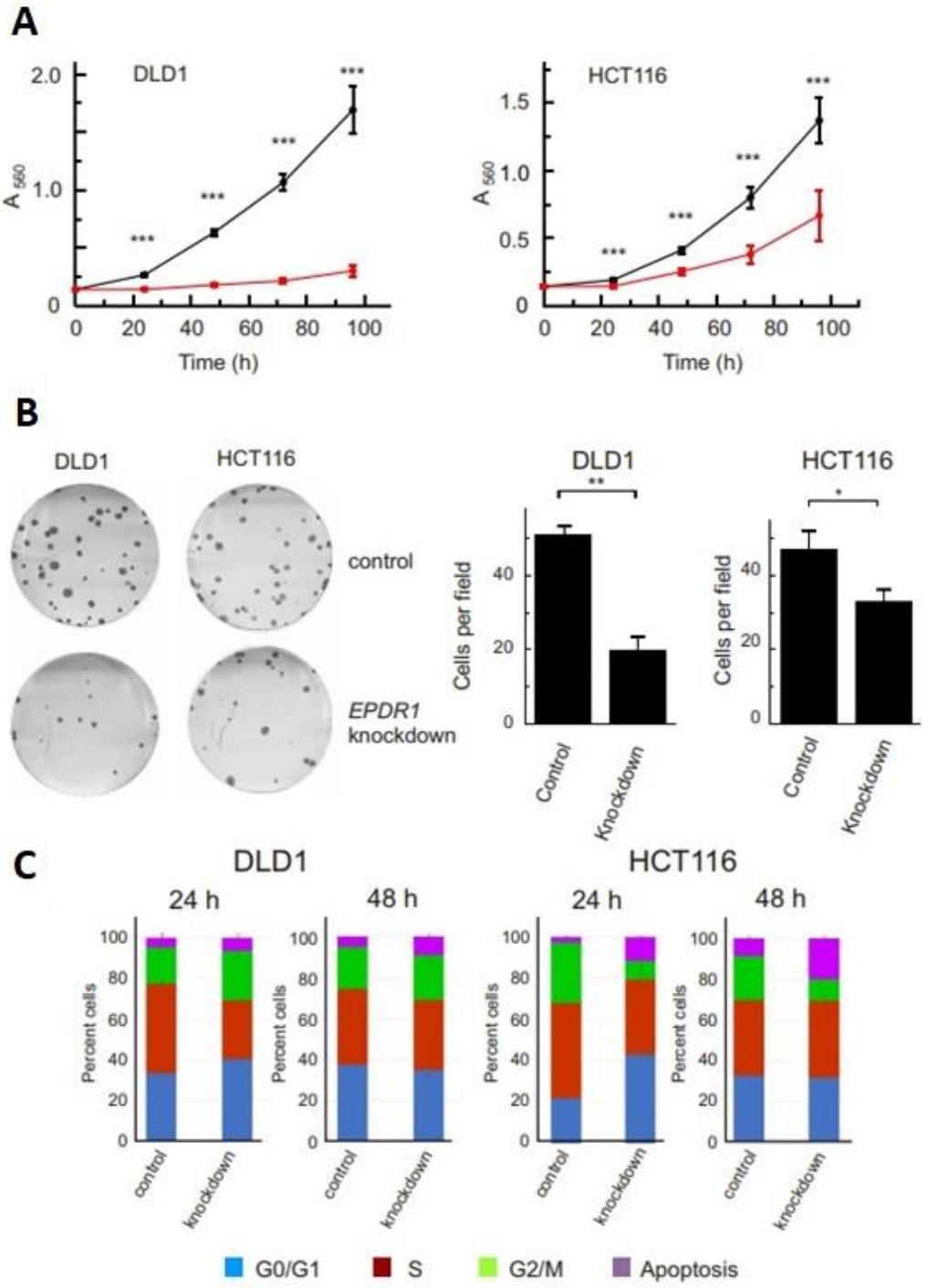

Figure 38: Effects of EPDR1 knocking-down with si2 on the growing of CRC cell lines. Cells were transfected with scrambled siRNA (control) or with si2 (EPDR1 knockdown) in DLD1 and HCT116 cell lines. A) Cell proliferation measured by MTT assay. The graph shows the results of an experiment, run in quintuplicate. The red lines correspond to the cells transfected with si2 while the black lines correspond to cells transfected with scrambled siRNA (control). B) Colony formation assays of cells treated with siRNA control and with si2. Both the photograph of a representative plate and the averaged quantification of 3 plates are given. C) Flow-cytometry cell cycle analysis of cells treated with siRNA control and with si2. *, $p < 0.05$; **, $p < 0.005$; ***, $p < 0.001$.

The effects observed in Figure 39 with the si2 corroborate the results obtained with si1 in relation to migration, invasion and adhesion properties of *EPDR1* silenced cells.

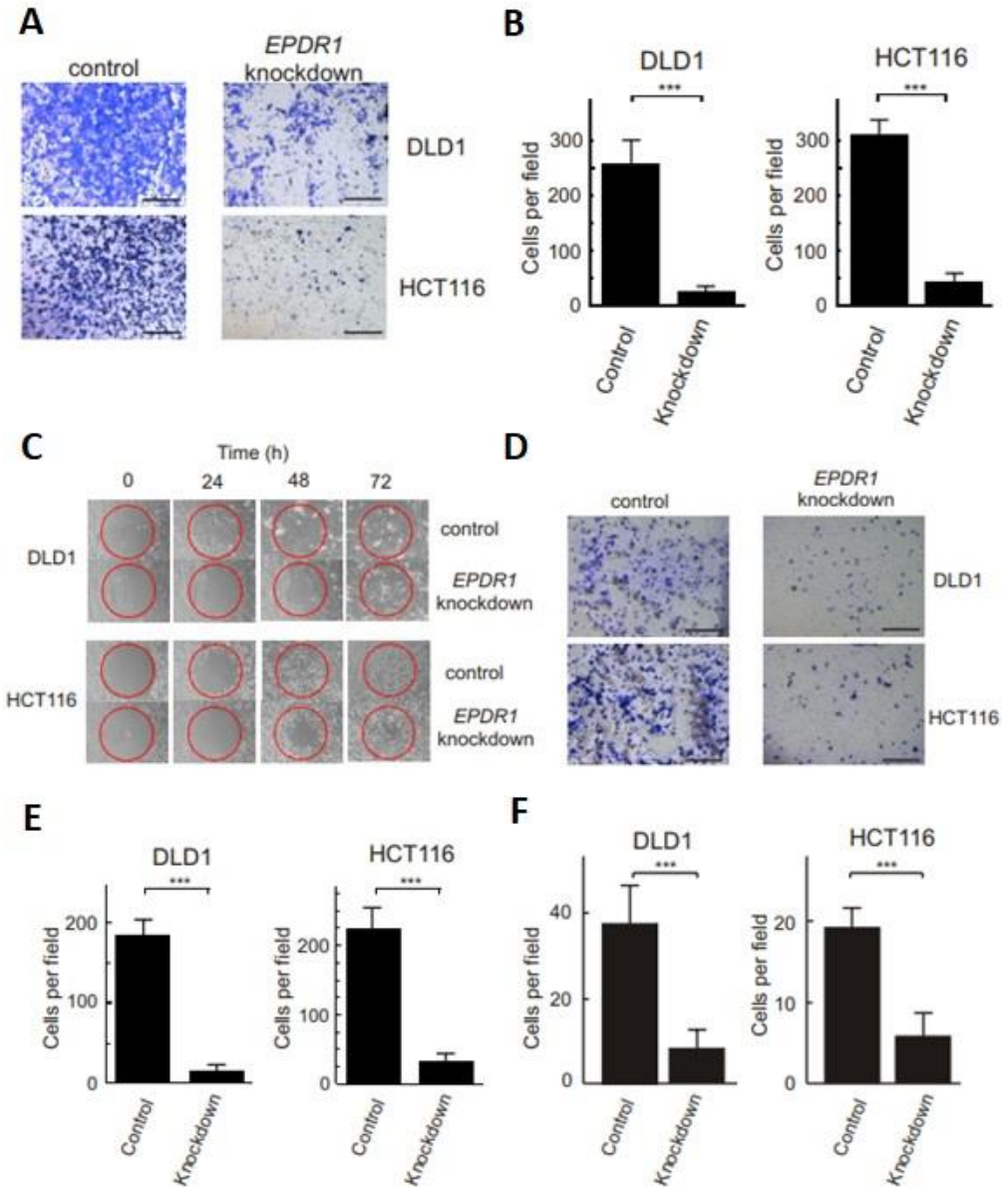

Figure 39: Effects of EPDR1 knocking-down with siRNA2 (si2) on the migration of CRC cell lines. A) Transwell migration assay of cells treated with scrambled siRNA (control) and with si2 (EPDR1 knockdown). A photograph of a representative plate is shown. B) Averaged quantification of the cell counting in 3 fields each from 3 transwell plates. C) Wound-healing assay of cells treated with scrambled siRNA (control) and with si2. The time after removing the central gel layer is shown. D) Assay carried out as in A, but through a Matrigel layer. A photograph of a representative plate is shown. E) Averaged quantification of the assay of 3 plates (3 fields each). F) Effects of knocking down the gene EPDR1 on the adhesion of cells to type I collagen-coated plates (average of 6 determinations). The size bars in A and D correspond to 200 µm.

4.1.2.5 Overexpression of EPDR1 gene using transitory plasmid transfection strategy in two cell lines

To further elucidate the role of *EPDR1* in CRC cell lines, SW48 and RKO cells, that do not express this gene[158], were transfected with plasmids encoding *EPDR1* (p*EPDR1*) and with control empty plasmid (pCAGGS), described in Material and Methods section.

To ensure the correct overexpression of *EPDR1* after transfection, an RT-qPCR experiment was made. As Figure 40 shows, in both, RKO and SW48 cells, there is a significant increase in the expression of *EPDR1*, which passed from being negligible or even undetectable to a level even higher than that of intact DLD1 and HCT116 cells (data not shown).

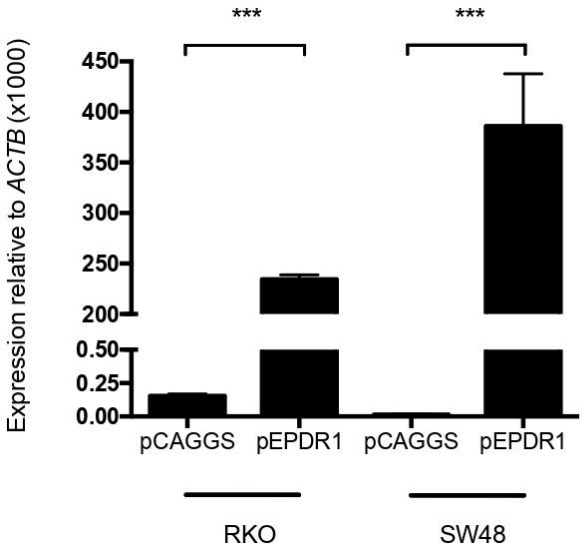

Figure 40: *EPDR1* overexpression in CRC cell lines. Transcript level of the gene relative to ACTB measured by RT-qPCR after 48 h of transfection with the empty plasmid pCAGGS or with the p*EPDR1* plasmid. The results were compared by Student's t-test. ***, p< 0.001.

The overexpression was also checked at a protein level as shown in Figure 41 for RKO cells. When cells were transfected with an empty plasmid, no specific staining was observed, while when the cells were transfected with p*EPDR1*, staining can be detected at the cellular membrane level and in some specific cytoplasmic areas. These results are compatible with the hypothetical structure due to the signal peptide and the possible function, based on our results of the *EPDR1* gene.

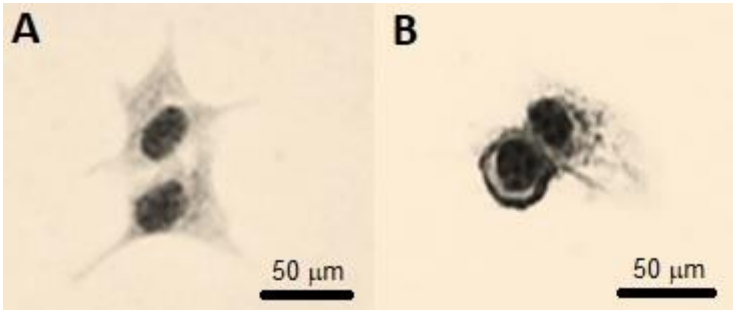

Figure 41: Detection by Immunocytochemistry of EPDR1 in CRC cell line RKO. A) Cells after 48 h of transfection with the empty plasmid (pCAGGS). B) Cells after 48h of transfection with p*EDPR1*. Staining was performed as indicated under Material and Methods.

4.1.2.6 Overexpression of *EPDR1 confers higher* cell dissemination and proliferation capacity in colorectal cancer cell lines

The effects of *EPDR1* knocking down with siRNAs on the phenotype of CRC cells have been described above. Here, the effects of overexpression on cell adhesion, migration, invasiveness and colony formation capacity are described to obtain a deeper understanding of the role of the gene in CRC progression.

Figure 42 shows that increasing the expression of *EPDR1* resulted in an increase in the binding capacity of both cell lines to type I collagen-coated plates.

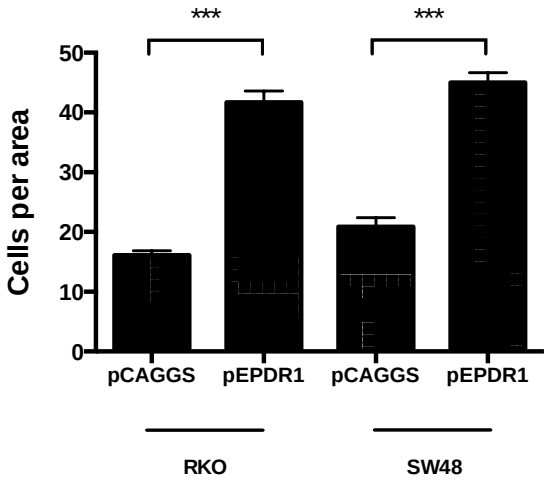

Figure 42: Effects of overexpressing *EPDR1* in RKO and SW48 on adhesion to type I collagen-coated plates. The averaged quantification of the assay of 3 plates ± SD are shown. Results were compared by Student's t-test. ***, $p < 0.001$.

The effects of *EPDR1* overexpression in migration and invasiveness of the cells were also studied by transwell analyses. A significant increase in the migration and invasion capacity of the RKO and SW48 cells was observed when these cell lines were transfected with p*EPDR1* (Figure 43). These results are in agreement with those obtained in the gene silencing experiment.

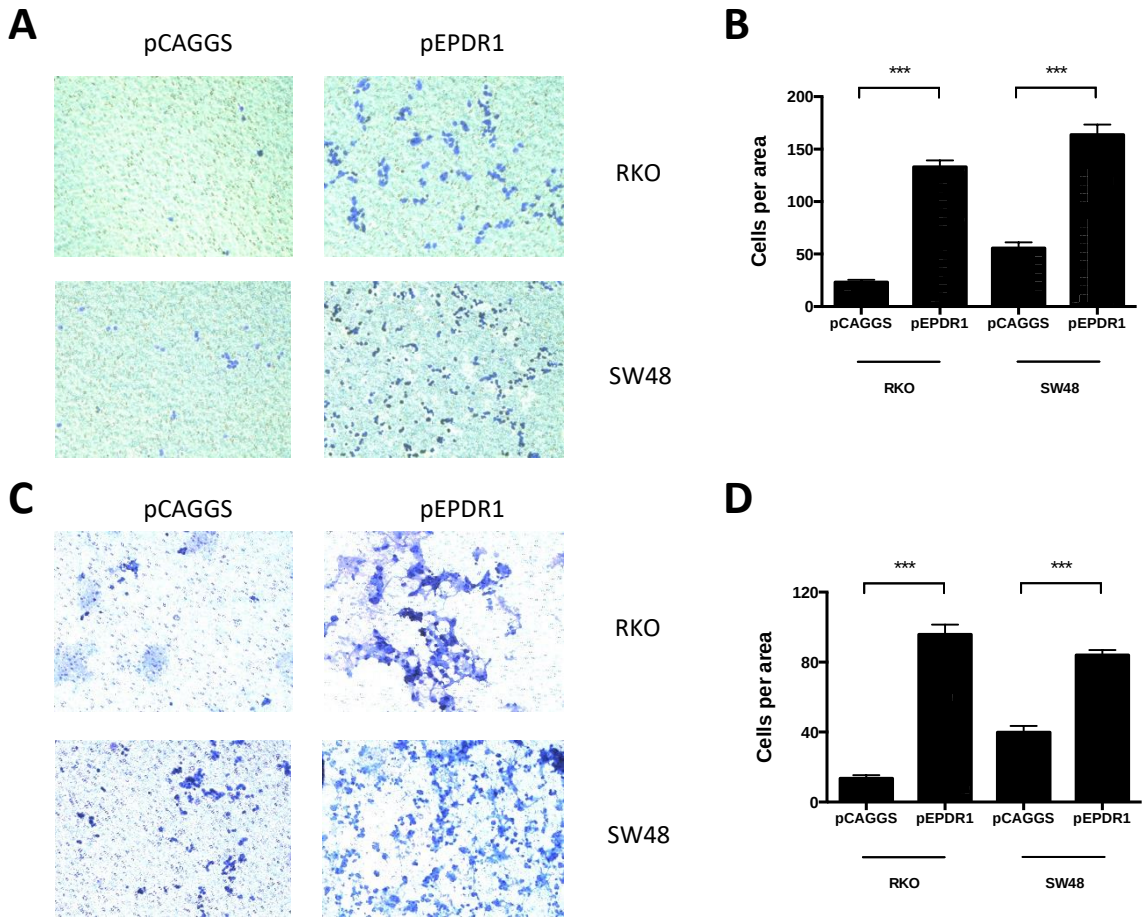

Figure 43: Effects of *EPDR1* overexpression on the migration and invasiveness of CRC cell lines. (A) Transwell migration assay of cells transfected with pCAGGs and with p*EPDR1*. Photographs of representative plates are shown. (B) Averaged quantification of the results of 3 transwell plates. (C) Assay carried out as in A, but through a Matrigel layer. Photographs of representative plates are shown. (D) Averaged quantification of the assay of 3 plates. The results were compared by Student's t-test. *** $p < 0.001$.

Overexpression of the *EPDR1* gene also had the effect of increasing the capacity of colony formation, as can be seen in Figure 44.

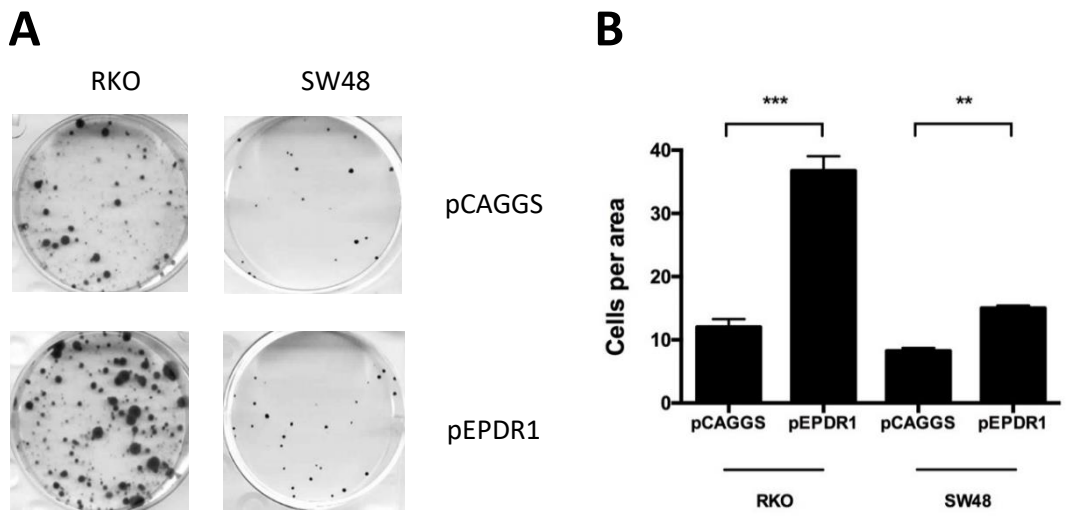

Figure 44: Colony formation test. The assays were carried out with cells transfected either pCAGGS or with p*EPDR1*. A) Photographs of representative plates. B) Averaged quantification of 3 fields in 3 plates. The results were compared by Student's t-test. ** $p < 0.005$; *** $p < 0.001$.

4.1.3 Regulation of the *EPDR1* expression in colorectal cancer

4.1.3.1 The expression of *EPDR1* is dependent on the methylation status of its promoter

It has been previously described that the expression of the whole *EPDR1* gene is not detected in RKO, almost negligible in SW48 (0.00012±0.00007 relative to *ACTB*), low in HCT116 (0.0054±0.0004), and high in DLD1 (0.0252±0.0056) and D-Mut1 (0.0454±0.0055)[158]. To explain the causes of this differential behaviour, it can be hypothesized that epigenetic changes may cause the differences in *EPDR1* expression among the cell lines used. To check that hypothesis, the methylation of a segment of 1241 bp, which includes a CpG island encompassing *EPDR1* exon 3, the first transcribed exon except for isoform 3, and the first translatable exon in the major isoform, was determined. A diagram of the location of the region to be studied can be observed in Figure 45.

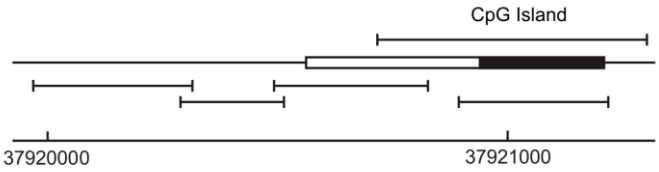

Figure 45: Scheme of the CpG island located at the beginning of exon 3 of *EPDR1*. The black region shows the translatable area. The location of the amplicons used for the study is shown. The scale gives the absolute position of the region, in bp, in chromosome 7.

The results confirmed the starting hypothesis, since the methylation pattern negatively correlates with the level of expression of *EPDR1* in the different CRC cell lines (Figure 46). In SW48 and RKO cells the level of methylation in the studied area was very high, and this may repress the expression of the *EDPR1* gene. However, in the case of HCT116, which shows an intermediate *EDPR1* expression value, the level of methylation was significantly lower than in the previous cell lines. Finally, the analysis of DLD1 and D-Mut1 revealed the lowest level of methylation, which correlates with the higher degree of expression of *EPDR1*.

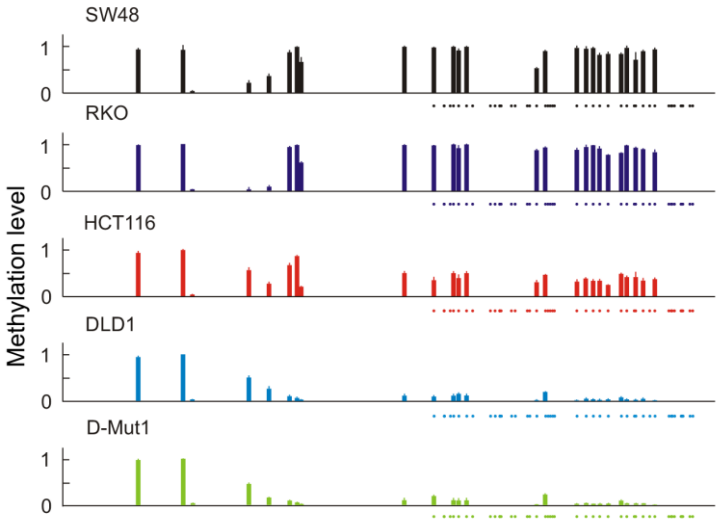

Figure 46: Quantification of CpG methylation in the region near the exon 3 promoter of *EPDR1*. The quantitative analysis was carried out using an AGENA's MassARRAY platform. The dots at the bottom of each chart mark the position of different CpG within the island, even when their methylation level could not be determined with the method used. The results represent the average of 5-10 determinations with their SD.

To demonstrate that the hypermethylation of the CpG island was involved in the repression of *EPDR1* in RKO and SW48 cell lines, the cells were treated with 5-azacytidine, a known demethylating agent. Figure 47 shows that there was a dose-dependent increase in *EPDR1* expression after 5-azacytidine treatment. Therefore, this experiment demonstrated the causal effect of DNA methylation on the expression of *EPDR1*.

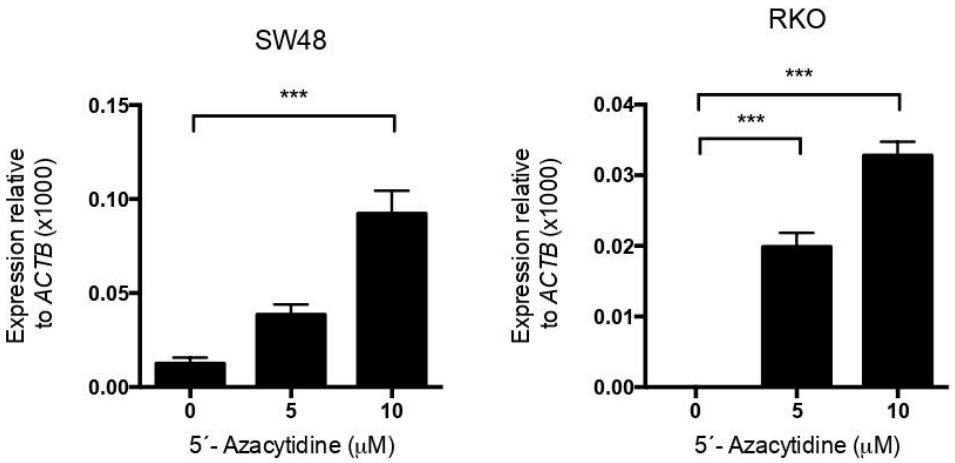

Figure 47: Effects of 5'-Azacitidine treatment on the expression of *EPDR1* in the cell lines SW48 and RKO. In both cell lines, a dose of 5 μM and a second dose of 10 μM of 5'-Azacitidine have been used. The results of *EPDR1* expression after treatment were assessed by RT-qPCR using the *β-actin* gene as housekeeping. The Kruskal-Wallis statistical test has been used to observe the differences between the different doses or with respect to no treatment. ***, $p < 0.001$

4.1.3.2 The expression of *EPDR1* is regulated by 193a-5p microRNA

It has been shown in the preceding section that DNA methylation regulates the expression of *EPDR1*. Nevertheless, some other regulatory mechanisms may coexist with that epigenetic modification. Taking into account that miRNAs are often involved in the regulation of gene expression, and that they play an important role in cancer,

the possible participation of these non-coding RNAs in *EPDR1* regulation was next examined.

It was observed by *in silico* analysis, with the TargetScan tool[195] that there was a direct and total interaction of the 193a-5p microRNA and the 5' end of exon 6 of *EDPR1*, common to the isoforms 1 and 2 of the gene. The binding complementarity is shown in Figure 48.

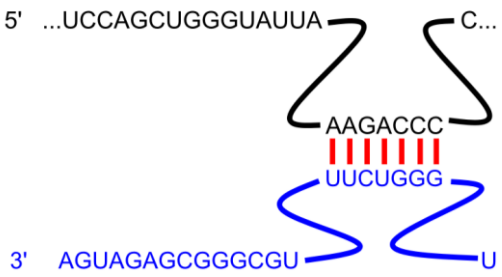

Figure 48: Sequence complementarity between EPDR1 and the microRNA 193a-5p. The sequence in black represents the *EPDR1* mRNA, while the blue sequence represents the 193a-5p microRNA. The red bars link the overlapping bases.

A correlation study was performed to assess the role of 193a-5p miRNA in the regulation of *EPDR1* expression in CRC cell lines that express or not this gene. As can be observed in Figure 49, the lines deficient in the expression of *EPDR1* gene showed a higher expression level of 193a-5p microRNA. The opposite was observed in the CRC cell lines that express *EPDR1*, such as DLD1 and HCT116. Of note, the difference in 193a-5p miRNA level between DLD1 and D-Mut1 cells may account for a fine tune regulatory mechanism that could explain the distinct levels of *EPDR1* transcription although the CpG island is virtually non-methylated in both cell lines.

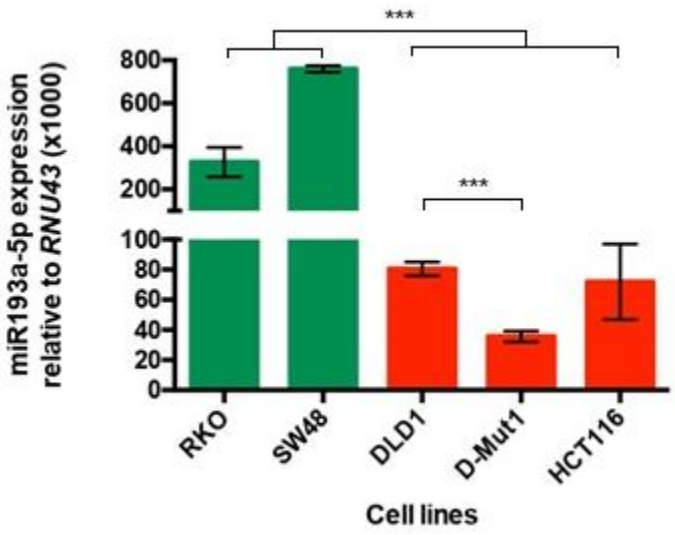

Figure 49: Expression level of microRNA 193a-5p relative to RNU43 determined by RT-qPCR in cell lines that express *EPDR1* (DLD1, D-Mut1 and HCT116) or not (RKO and SW48). For the comparison of the expression of miR-193a-5p in the different cell lines a Kruskal-Wallis test has been used. ***, $p < 0.001$.

To further analyse whether a functional relationship exist between the microRNA 193a-5p and *EPDR*1, cells were transfected with a mimic of 193a-5p microRNA. At first sight, it may be surprising that, contrary to what was expected, overexpressing the mimic, resulted in an increase in the expression of *EPDR1* after 24h and 48h of transfection (Figure 50). As it will be further commented under the Discussion, the observed increase in the *EPDR1* expression, especially after 48 h after transfecting with the mimic (Figure 50, A), may be an artefact resulting from the dramatic overexpression of the mimic (Figure 50, B).

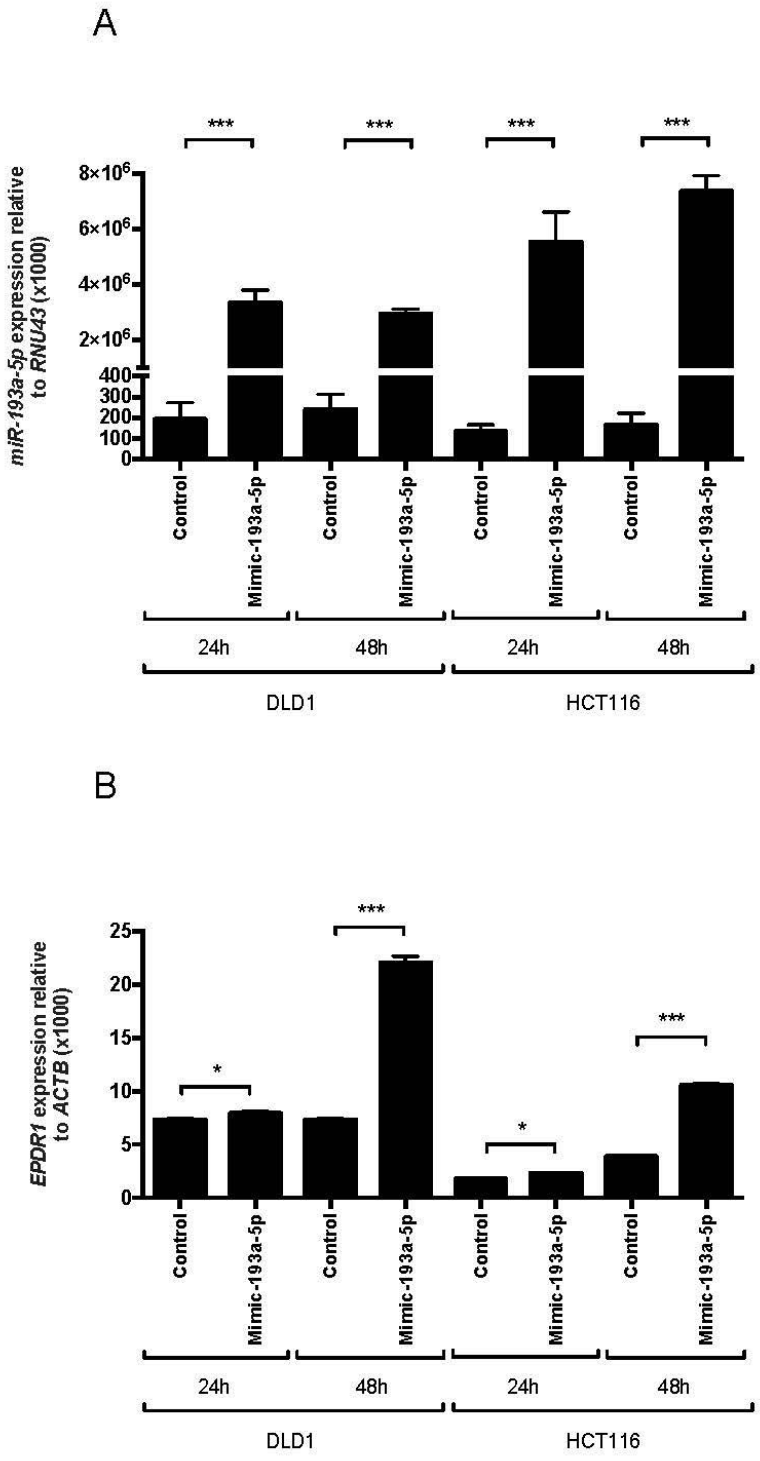

Figure 50: Expression of miR-193a-5p and *EDPR1* gene determined by RT-qPCR at 24 and 48 h after transfection. The valuation of the expression has been analysed in DLD1 and in HCT116. A) Expression of miR-193a-5p relative to *RNU43* as housekeeping gene. Notice the scale cut on the ordinate axis. B) Expression of *EPDR1* relative to β-actin as housekeeping gene. A Student´s t-test has been used to assess the difference in expression of both miR-193a-5p and *EPDR1*, in the presence of the mimic or in its absence (control). *, $p < 0.05$; ***, $p < 0.001$.

To analyse if there is a direct effect of miRNA 193a-5p over EPDR1 transcript, avoiding secondary effects of mimic over endogenous gene, RKO cells were double transfected, with 193a-5p mimic and the overexpression plasmid p*EPDR1*. As a control, cells were transfected with mimic and empty control plasmid.

Intact RKO cells expressed very low levels of *EPDR1* (Figure 51) and when the cells were transfected with the 193a-5p mimic, the expression of the endogenous gene increases slightly. Interestingly, when the cells were also transfected with p*EPDR1* to increase the levels of expression of the gene, the treatment with the 193a-5p mimic does decrease the expression of *EPDR1* in a significant manner.

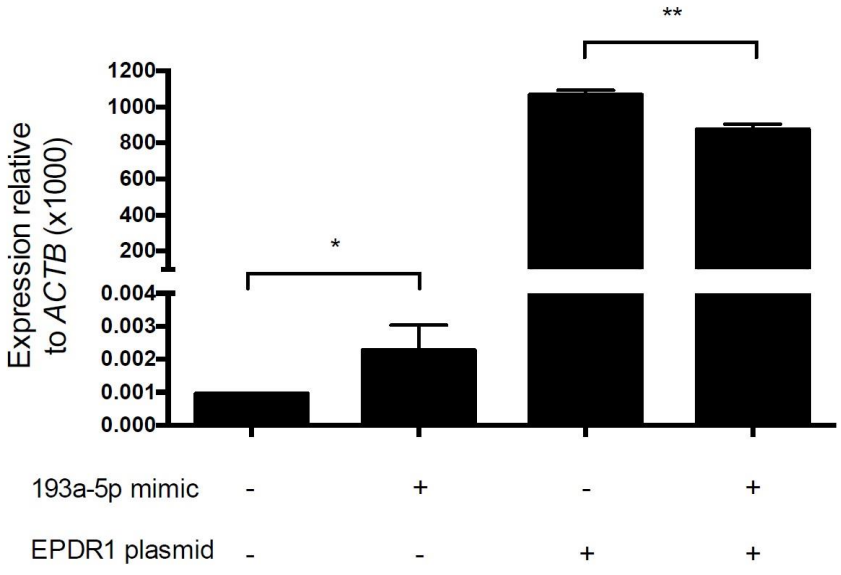

Figure 51: Effects of transfecting RKO cells with p*EPDR1* and/or with 193a-5p mimic on the expression of *EPDR1* determined by RT-qPCR. β-actin was used as a housekeeping gene. Note the change of scale in the axis of the ordinate. A Student´s t-test was used to assess the expression after transfections. *, p <0.05; **, p <0.01

Once determined that 193a-5p miRNA could regulated in some manner the expression of *EPDR1* gene, a study was performed to examined if there was a correlation between the expression of *EPDR1*, and microRNA 193a-5p in our patients

cohort. This study was carried out with the 77 samples of paired fresh-frozen tissues from the HCV patients' cohort described above. It can be seen in Figure 52, that there is a significant negative correlation (p-value =0.0006) between the expression of microRNA 193a-5p and *EPDR1*. In other words, an increase in the expression of microRNA 193a-5p significantly reduces the expression of *EPDR1*. Therefore, it is possible that the regulation of *EPDR1* expression by that microRNA, observed *in vitro* also exists *in vivo*.

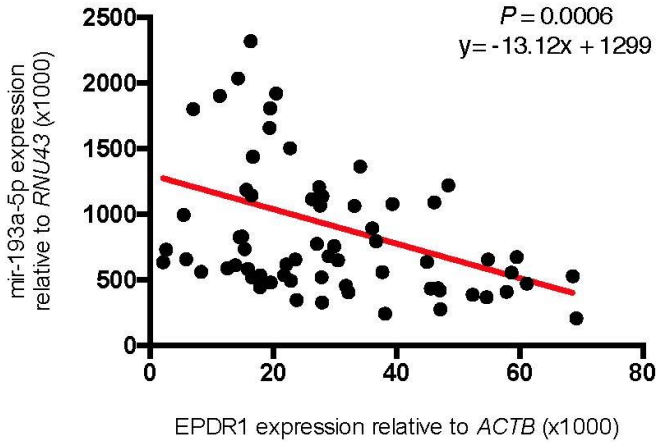

Figure 52: Correlation between the expression of 193a-5p microRNA and *EPDR1* in the CRC patients' cohort. To obtain correlation data between both parameters, a linear regression has been performed. In addition, the formula of the straight line (in red) has been calculated by means of data interpolation, being able to develop a predictive model of expression in the population in a significant way (p-value = 0.0006).

4.1.3.3 The level of *EPDR1* expression relates to the mutational status of colorectal cancer patients

It was reported in a previous work from our laboratory that the expression of *EPDR1* is correlated to the mutational status of *KRAS* in cell lines[158]. Taking advantage of the possibility of experimentally determine the mutational profile of CRC samples, it has been possible to relate the expression level of *EDPR1* with *KRAS* and other CRC-

related mutations detected in patients of the HCV cohort. To simplify the analysis, it was established an expression score of *EPDR1* based on the average of gene expression in healthy samples. So, values above such average were considered as high. Results were represented graphically in Figure 53. Among the patients with high expression, only 3% had mutations in *CTNBB1* and *BRAF* and 13% of the patients had mutations in *TP53*. Interestingly, mutations in *KRAS* or *APC* together, account for 57% of the mutations present in patients with high *EPDR1 expression*. On the contrary, patients classified as low expression of *EPDR1* present different mutation profiles, consisting of 30% of mutations in *TP53,* 15% of mutations in *PIK3CA*, 12% of mutations in *BRAF*, 8% of mutations in *NRAS and 4% of mutations in FBXW7*. In summary, it is important to notice that *KRAS* and *APC* mutated patients were mainly found in the high *EPDR1* expression group while in the low expressing one there was an enrichment of *TP53* and *BRAF* mutated (see the Discussion section).

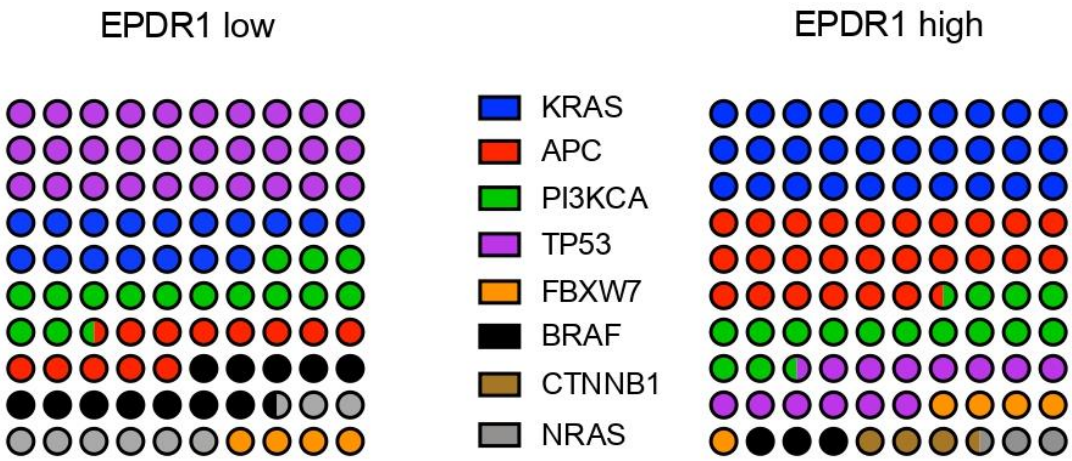

Figure 53: Graphic representation of the mutations percentage presents in the patients with CRC from HCV cohort based on the *EPDR1* expression. The sample size of the cohort that has undergone a mutational determination is 83. On the left, the percentage of mutations present in patients with an *EPDR1* value lower than the mean is shown. On the contrary, the distribution of the mutations presented in patients with high expression of *EPDR1*, that is, above the mean, is shown on the right. Each point of the panels represents one unit of a total of 100.

The beeswarm plot of Figure 54 shows that a positive, significant correlation exists between the distribution of patients in high- or low- *EPDR1* expression groups and the presence of *KRAS* or *APC* mutations. On the contrary, the correlation with *BRAF* mutations is negative. This result suggests that the signals transduced by the *KRAS* and/or *WNT* pathways may be relevant for *EPDR1* expression.

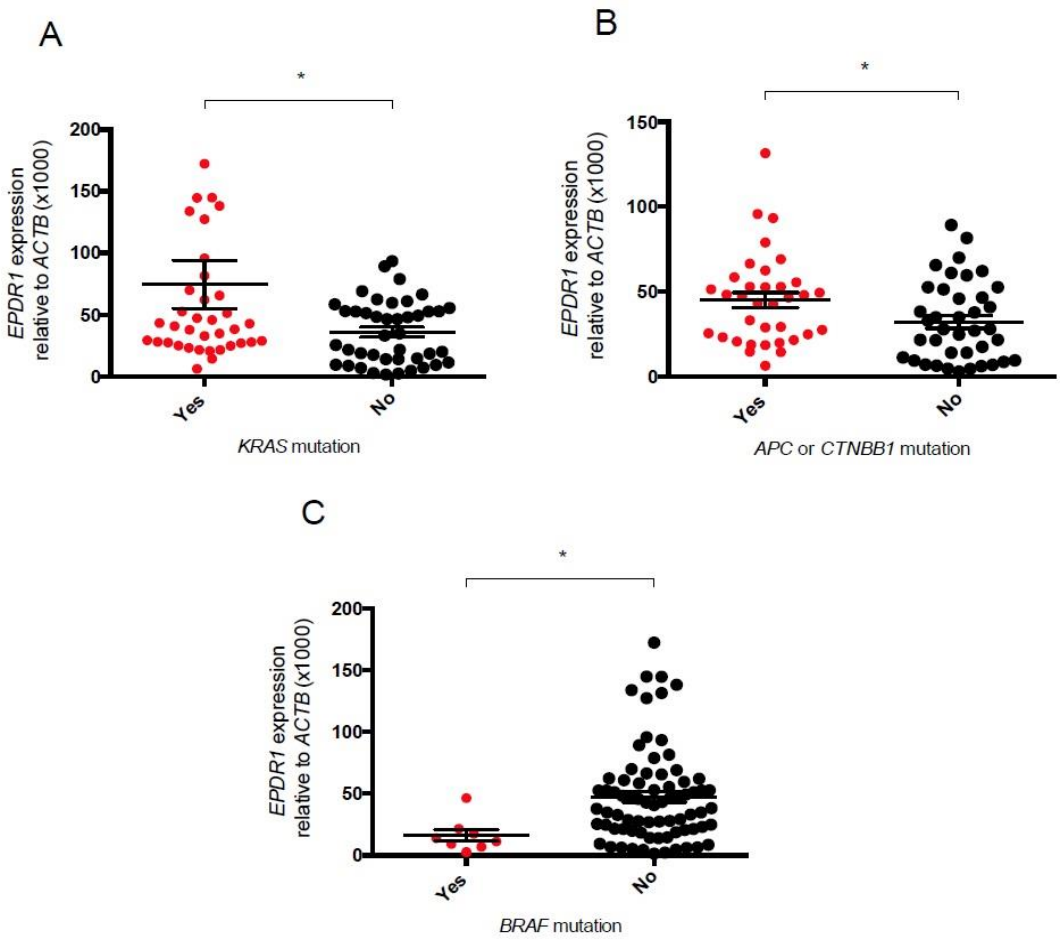

Figure 54: Beeswarm plot representing the expression of *EPDR1* in relation with the mutational status of CRC patients. *EPDR1* expression relative to ACTB related to (A) the presence or not of *KRAS* mutation (B) the presence of *APC* or *CTNBB1* mutations and (C) the presence of *BRAF* mutation. The results were compared by Student's t-test. *, $p < 0.05$.

Thanks to Dr. Tarazona our HCV cohort of patients is well classified according to the Consensus Molecular Classification (CMS) of CRC. In this manner, it could be investigated whether there was a correlation between the expression of *EPDR1* and

the CMS of our cohort. As shown in Figure 55, the patients harbouring higher expression of *EPDR1* were classified in the CMS3 subtype. Patients classified as CMS3 are characterized by having mutations in *KRAS* and *APC*, among other alterations. According to the correlation of *EPDR1* expression with the mutation profile, it can be observed that those patients with low *EPDR1* expression levels were mainly included in the CMS1 subgroup, characterized by the prevalence of *BRAF* mutations. On the other hand, CMS1 subtype is defined by encompassing patients with a hypermethylated phenotype in contrast with CMS3 subtype including hypomethylated subjects. These findings are in concordance with the expression and regulation of *EPDR1*. It is worth highlighting the significant *EPDR1* expression level found in patients classified into CMS4 subgroup. CMS4 tumors are characterized by activation of pathways related to EMT activation, fact that is in agreement with present *in vitro* functional studies of EPDR1 in CRC cell lines. Likewise, the distribution of our cohort according to CMS subtypes is in line with our hypothesis.

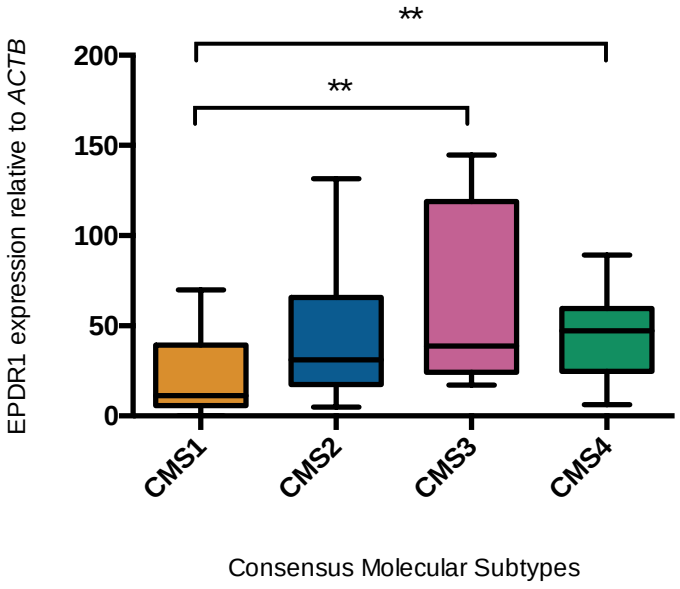

Figure 55: Box plot representation of the *EPDR1* expression in the patients from the HCV cohort classified into the four CMS subgroups. The sample size in each group was 22 patients in CMS1, 22 patients in CMS2, 17 in CMS3 and 25 in CMS4. For the comparison between the CMSs the Kruskal Wallis test has been used. **, $p < 0.01$.

4.2 Study of *ZNF518B* gene as a putative biomarker in colorectal cancer progression

4.2.1 *ZNF518B* gene expression in patient samples

4.2.1.1 The *ZNF518B* gene and its isoforms are overexpressed in tumour samples of colorectal cancer with respect to healthy samples

ZNF518B expression in patient samples was determined both in the TissueScan cDNA array (OriGene) and in the samples from the HCV patient cohort as it has been described above for the gene *EPDR1*. The relevant clinico-pathological characteristics relative to both collections of samples have been described under Materials and Methods. In the OriGene cDNA array, the expression levels of the global gene (Figure 56, panels A and B) and of its isoforms 1 and 2 (Figure 56, panels C and D) were determined by RT-qPCR. Both, the whole gene and its isoforms are significantly overexpressed (p-value < 0.05) in human CRC at all stages but no significant differences were observed from one stage to another.

The expression of the *ZNF518B* gene was next studied in samples from the HCV cohort. The overall comparison in tumour samples and non-tumour samples shows that there exists a significant difference (p-value < 0.001) (Figure 57). As the total expression of the *ZNF518B* gene was relatively low, in many cases the results were below the detection threshold of the qPCR, and only a fraction of the samples of the cohort could be analysed.

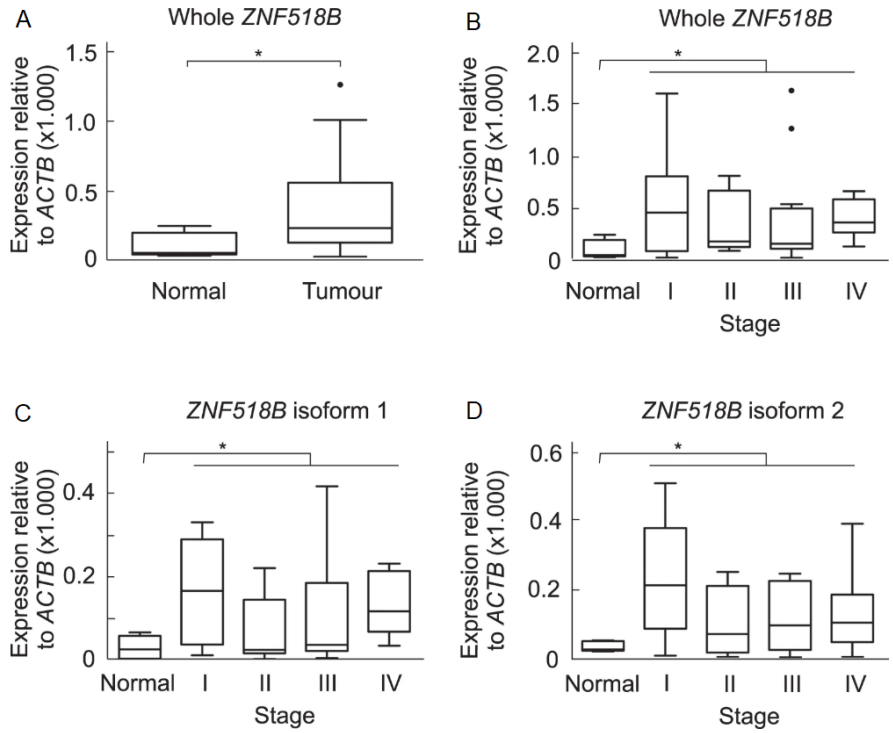

Figure 56: Expression of the *ZNF518B* gene relative to *ACTB* in the TissueScan cDNA array (Origene). The figure shows box plots with whiskers with maximum 1.5 interquartile range of the *ZNF518B* expression. The plots correspond to the comparison of whole gene and its major isoforms in normal *versus* all tumour tissues or stage-classified tumour tissues. Results were compared with the Kruskal-Wallis test. A) Expression of the whole gene in all the tumour samples. B) Expression of the whole gene with respect to stages. C) Expression of isoform 1 in different stages of disease development. D) Expression of isoform 2 in different stages of disease development. *, $p < 0.05$.

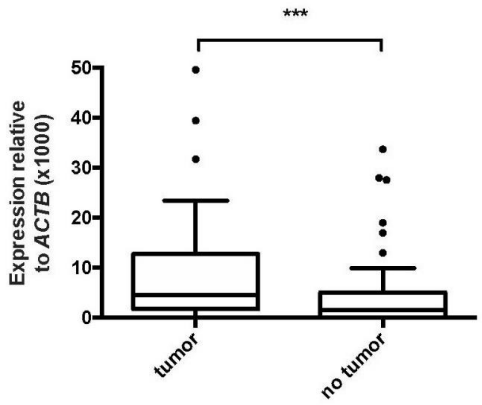

Figure 57: Box plot representation of the *ZNF518B* gene expression in the patient's cohort measure by RT-qPCR using *ACTB* as reference. The expression of total gene in normal and non-tumour adjacent tissues is represented. The results were compared by Student's t-test. ***, $p < 0.001$.

The quality of the RNA isolated from paraffin-embedded samples was not good enough to separately analyse the expression of the two main isoforms of *ZNF518B,* whose expression is low. To do this, it was mandatory to use frozen fresh tissues but, as previously mentioned, only 77 tumour samples of the cohort were available in that conditions. Results shown in Figure 58 revealed that isoform 2 is significantly more expressed than isoform 1 in this cohort of patients (p-value <0.001).

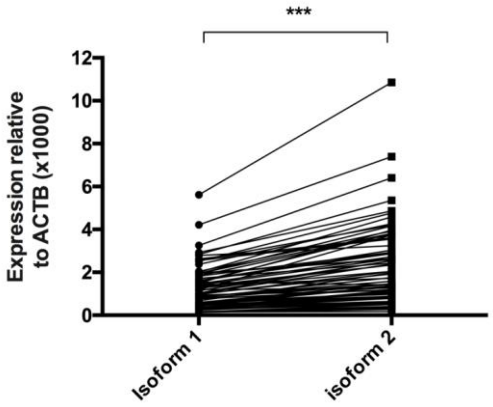

Figure 58: Expression of isoforms 1 and 2 of the *ZNF518B* gene in the cohort of 77 fresh samples from HCV CRC patients. The results were compared by T-test of paired sample. ***, p < 0.001.

The *ZNF518B* is a gene with unknown function but it seems clear that isoforms 1 and 2 probably act in different processes. Indeed, it would be suggested that isoform 2 could compete with isoform 1 function as a transcription factor. Isoform 2 retains the capacity to interact with other proteins, as, for example, the methyltransferase G9A, but lacks the possibility to bind DNA due to the absence of zinc fingers. Then, changes in the isoform 2/isoform 1 ratio could be a clinically interesting parameter to study. Both isoforms were analysing in each of the fresh samples and the ratio isoform2/isoform1 was correlated with different clinico-pathological parameters. Ratio changes do not correlate with most of them, except with the occurrence of disease relapse, as can be seen in Figure 59. Actually, the relapse in the disease is only observed in those patients with a lower ratio isoform 2/isoform 1. Despite of it, a Cox regression and Kaplan-Meier were done, without obtaining significant results.

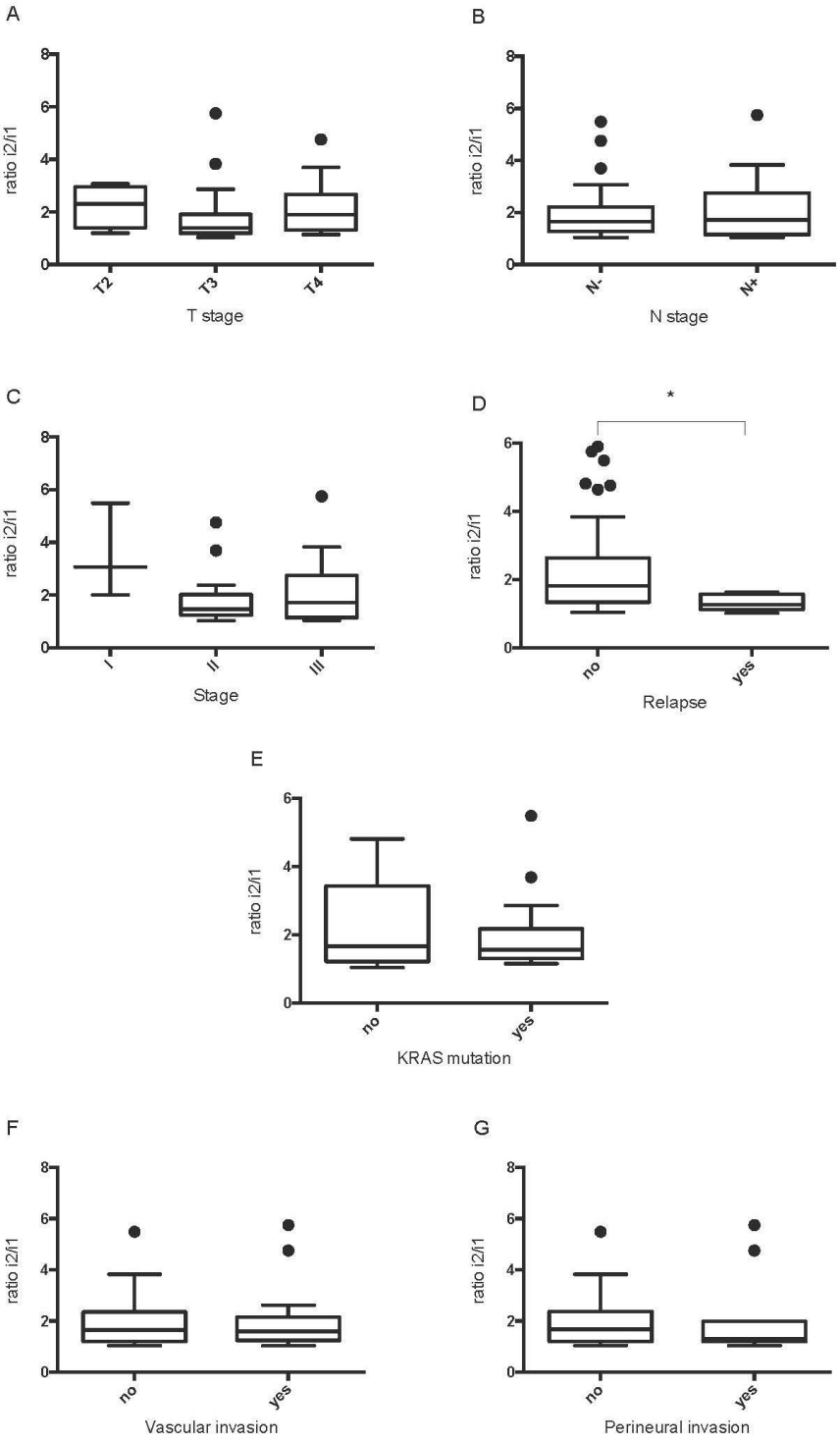

Figure 59: Representation of the ratio of expression between isoform 2 and isoform 1 of *ZNF518B* with respect to different parameters. The analysis was carried out with the fresh tissue samples of the cohort. A) Parameter T. B) Parameter N. C) Representation of stages I, II and III of CRC. D) Absence or presence of relapse in the disease. E) Presence or absence of mutation in *KRAS* in the cohort of fresh samples F) Vascular invasion in patients. G) Perineural invasion in patients. In A and C panels a Kruskal-Wallis test was performed, in rest of the panels a T-test statics was used. *, $p < 0.05$.

4.2.2 Functional analysis of the ZNF518B gene in colorectal cancer cell line models

4.2.2.1 Knockdown of *ZNF518B* gene using small RNA interference (siRNAs)

Following the same strategies described above for studying the role of *EPDR1* in CRC progression, to explore the function of the gene *ZNF518B* silencing assays were first performed. In a previous work, our group demonstrated that DLD1 and HCT116 CRC cell lines, express *ZNF518B*[158], so they were chosen to be transfected by siRNAs. After checking by RT-qPCR the effects of several commercial siRNAs on the knocking-down of both major isoforms of the gene, two of them were further used either as a pooled equimolar mixture or individually. As shown in Figure 60, the efficiency of the siRNA pool is more efficient for isoform 1 than for isoform 2 of the gene but in all cases the efficiency of the silencing was significant (p-value <0.001).

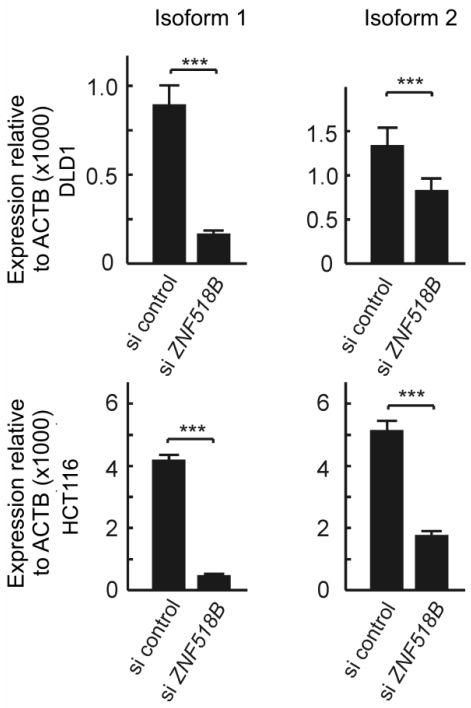

Figure 60: Expression of *ZNF518B* in isoforms 1 and 2 48 h after transfection of DLD1 and HCT116 with pooled siRNAs. The results were compared by Student's t-test. ***, p < 0.001.

Figure 61 shows the efficiency of the silencing of the two siRNAs of *ZNF518B* separately. It was decided to use siRNAs 1 and 2 as a mixture for knocking-down experiments since the silencing efficiency was improved respect to the individually effect.

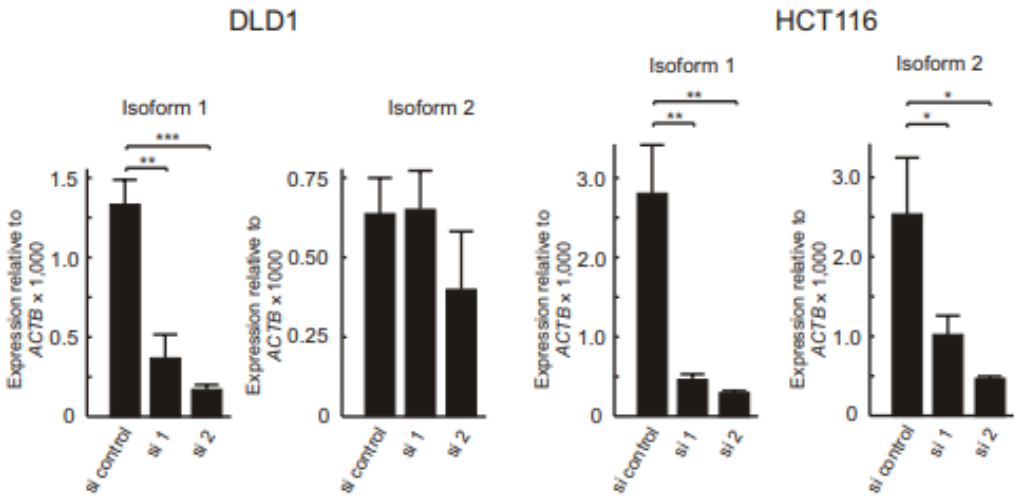

Figure 61: Efficiency of individual siRNAs, after 48 h posttransfection, on the expression of *ZNF518B* isoforms. The knocking down level was quantified in DLD1 and HCT116 cell lines by RT-qPCR. The results were compared by Student's t-test. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

4.2.2.2 Knockdown of *ZNF518B* gene expression revealed non-significant effects on cell proliferation in colorectal cancer cell lines

The consequences of the knocking-down in cell proliferation were evaluated by MTT, cell cycle and colony formation assays. The MTT assay revealed that there was no appreciable influence on proliferation in DLD1, although a decrease in the expression of the gene causes a small, but significant, reduction in the proliferation of HCT116 cells after 96 hours of silencing of *ZNF518B* (p-value <0.001) (Figure 62). As can be seen in Figure 63 the efficiency of the silencing at 96 hours is maintained. In accordance with these results, cell cycle analysis revealed no significant differences in any of the cell lines after *ZNF518B* knocking-down (Figure 64).

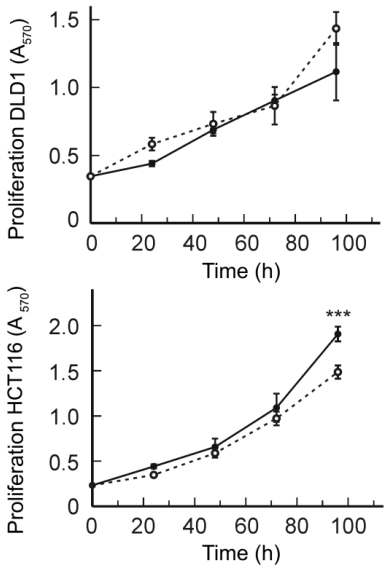

Figure 62: Effects of *ZNF518B* silencing on cell proliferation measured by MTT test in two *ZNF518B*-expressing cell lines, DLD1 (top) and HCT116 (bottom). The graph shows the results of two independent experiments, run in triplicate. The continuous line corresponds to the control cells transfected with scrambled siRNA, while the discontinuous line corresponds to *ZNF518B* knockdown cells. The results were compared by Student's t-test. ***, $p < 0.001$.

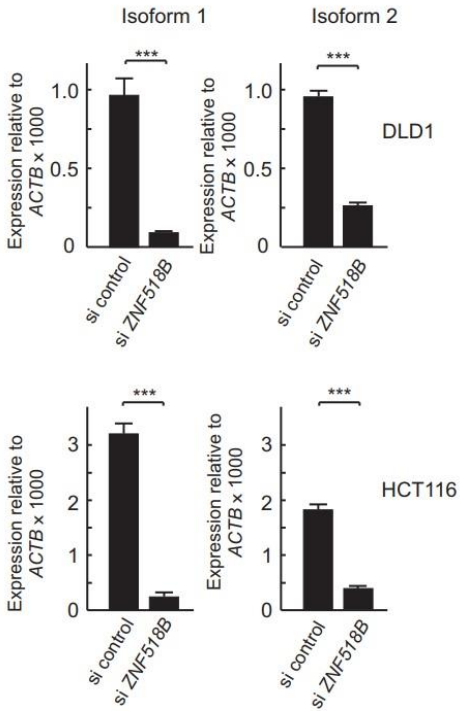

Figure 63: Efficiency of knocking-down for isoforms 1 and 2 of *ZNF518B* in DLD1 and HCT116 cells after 96h of transfection with scrambled or with the mixture of siRNAs. The results were compared by Student's t-test. ***, $p < 0.001$.

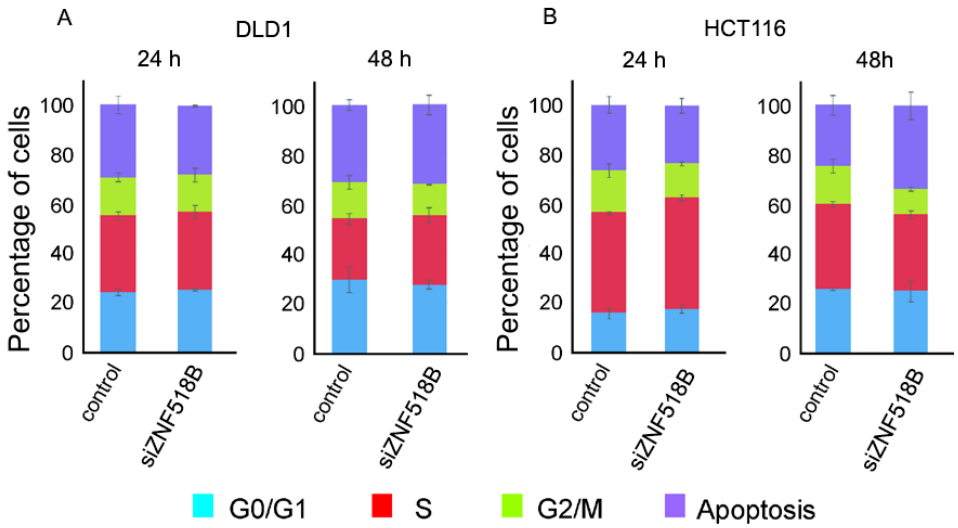

Figure 64: Study of the cell cycle of the CRC cell lines after silencing the *ZNF518B* gene. The distribution of cell cycle phases was determined at 24 and 48 h in DLD1 cell line (A) and in HCT116 cell line (B).

As a complement of the above analyses, the results obtained in the colony formation assay (Figure 65) demonstrated that the *ZNF518B* factor is not crucial in the proliferation capacity or in the anoikis phenomenon in the DLD1 and HCT116 lines.

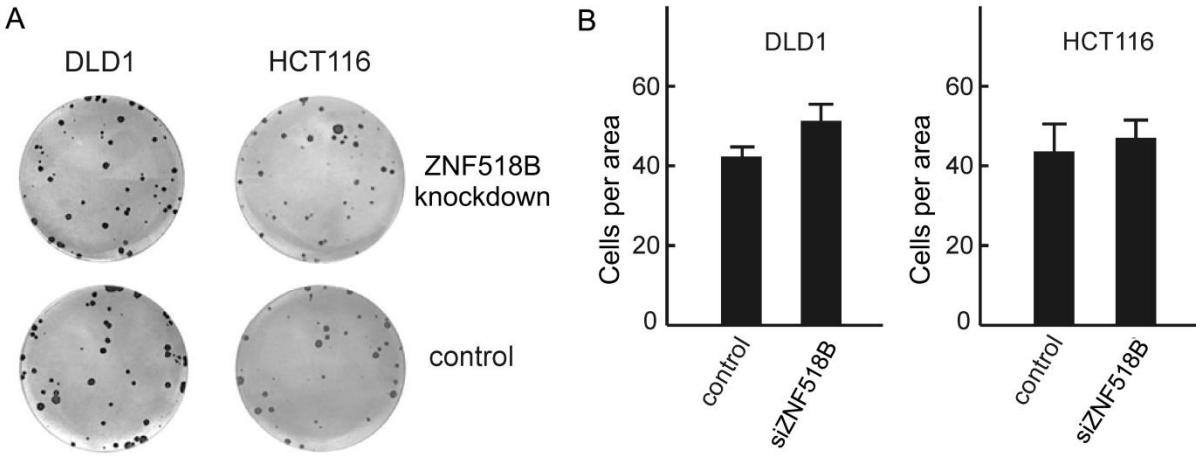

Figure 65: Effects of *ZNF518B* knockingdown in colony formation capacity. A) Photographs of representative plates are shown. B) Average quantification of the results of 3 different plates. The results were compared by Student's t-test.

4.2.2.3 Role of *ZNF518B* in favouring the epithelial-mesenchymal transition.

To evaluate whether ZNF518B is involved in the migration ability of CRC cells, a gene silencing with siRNAs was performed in DLD1 and HCT116 as a first strategy. After that, the *ZNF518B* knockingdown effects were studied by transwell analysis and by wound healing assays. The results of the transwell analysis (Figure 66) show that *ZNF518B* knocking-down significantly reduces cell migration (p-value < 0.001). The migration assays were also done with the individual siRNAs and in both cases a significant reduction of cell migration after knocking down the gene was observed (Figure 67). These results show that the two siRNAs do not have off-target effects when they are in combination.

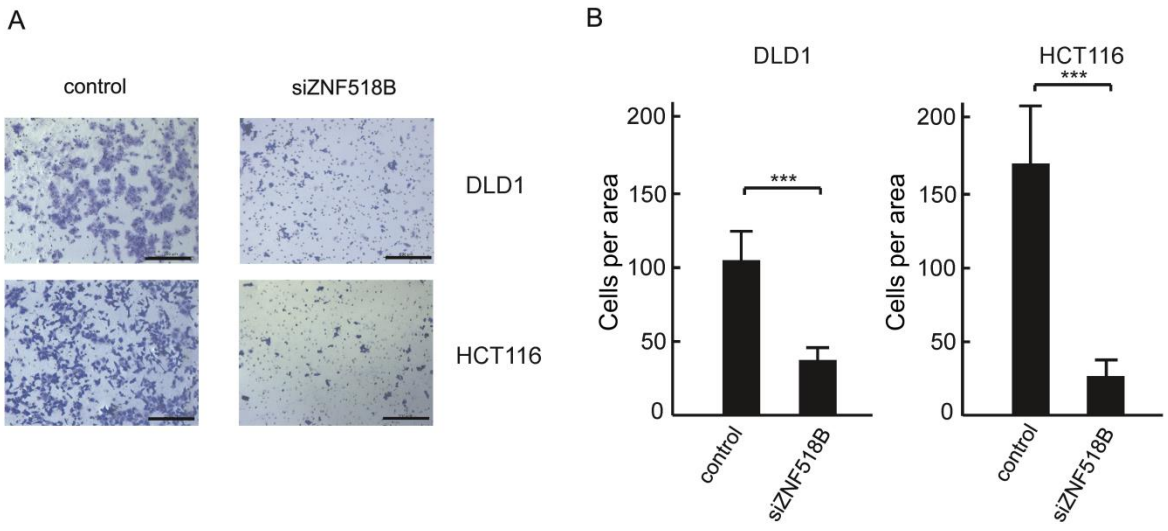

Figure 66: Transwell migration assay after silencing *ZNF518B*. A) Representative images showing the suppression of migration of DLD1 and HCT116 cells by knocking-down the *ZNF518B* gene. A photograph of representative plates is shown. The size bars correspond to 200 μm. B) Quantitative analysis showing the averaged quantification of 3 fields in 3 plates each. The results were compared by Student's t-test. ***, p < 0.001.

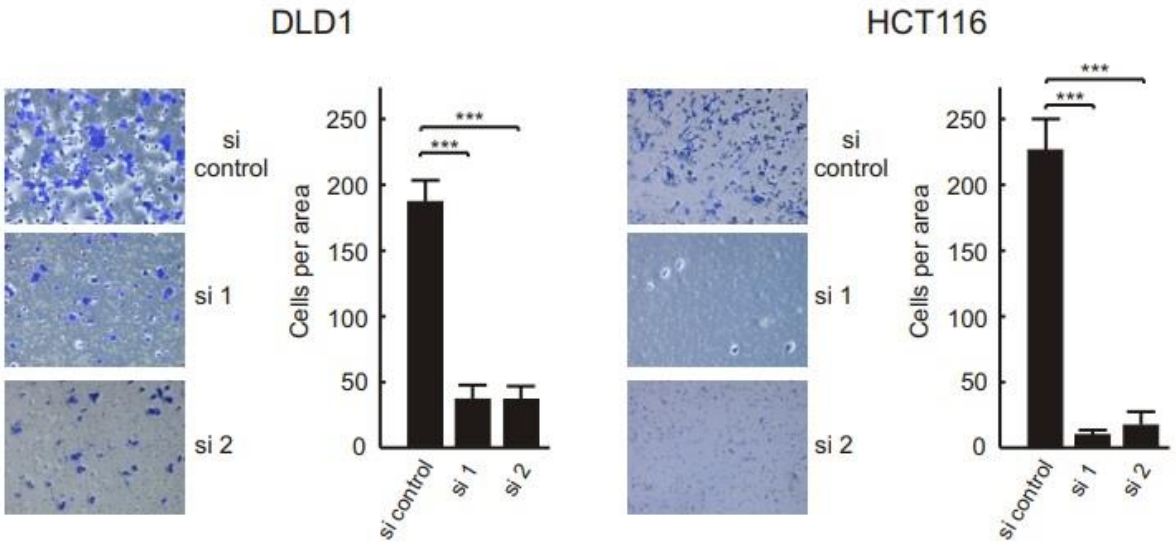

Figure 67: Effects of individual *ZNF518B* siRNAs on the suppression of migration of DLD1 and HCT116 cells in transwell assays. Photographs of representative plates are shown. Quantitative analysis shows the averaged quantification of 3 fields in 3 plates each. The results were compared by Student's t-test. ***, $p < 0.001$.

The wound healing assays confirmed the role of *ZNF518B* in cell migration ability. Significant differences (p-value <0.001) were observed in both CRC cell lines when *ZNF518B* silenced cells were compared with controls, as shown in Figure 68. At 24 hours no appreciable differences were observed between the experimental group and the control in either of the two cell lines. Nevertheless, differences were noticeable at 48 hours, with the gap practically closed in control cells compare with *ZNF518B* silenced cells.

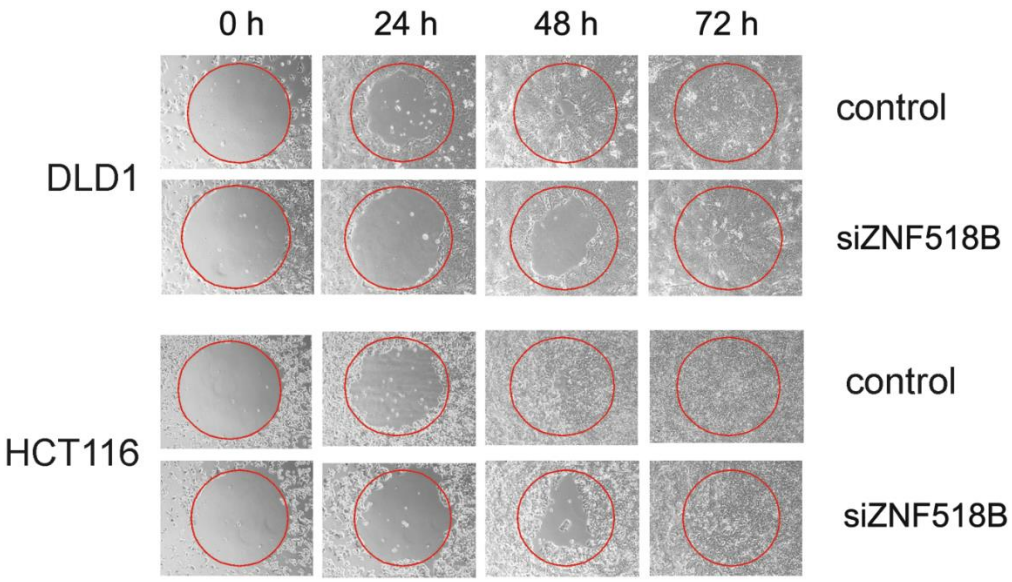

Figure 68: Wound healing assay showing the diminution of cell migration after silencing the *ZNF518B* gene (si*ZNF518B*) in cell lines DLD1 and HCT116. The red circle indicates the starting position of the gap at 0 h. The time of culture after removing the central gel layer is given at the top.

As indicated under Material and Methods, Matrigel, was used as an extracellular matrix model, and so the invasive capacity of the cells can be assessed. In this context, it should be highlighted that when the transwell analysis was carried out through a Matrigel layer, it was observed that silencing of the gene significantly reduces the invasion of cells through extracellular matrix (Figure 69). Therefore, it can be concluded that *ZNF518B* could have a role in the invasive capacity of DLD1 and HCT116 cell lines.

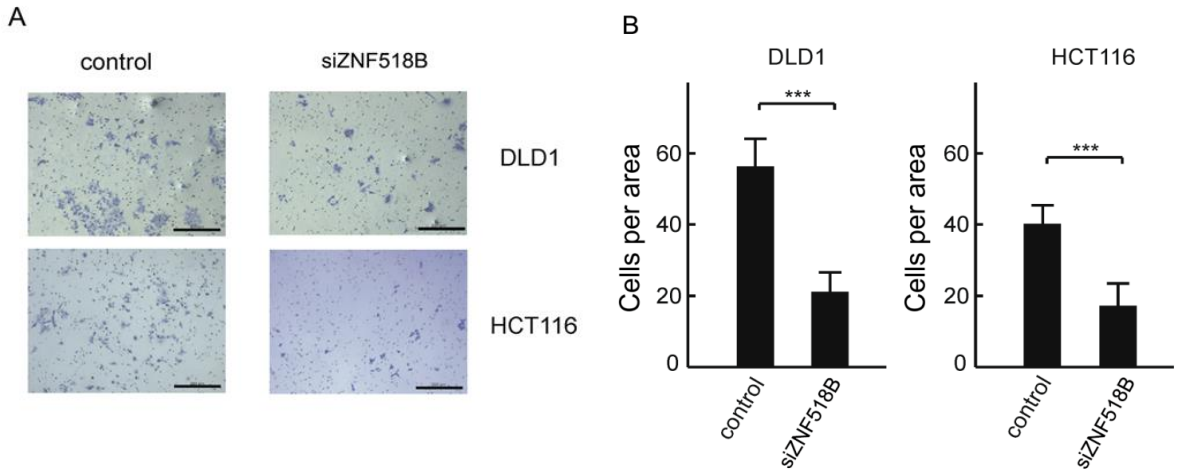

Figure 69: Effects of knocking down *ZNF518B* expression on invasion cell capacity. A) Transwell migration assay through Matrigel layer of cells treated with scrambled siRNA (control) and with siZNF518B (*ZNF518B* knockdown). A photograph of representative plates is shown. The size bars correspond to 200 µm. B) Averaged quantification of the results of 3 transwell plates. The results were compared by Student's t-test. ***, $p < 0.001$

The preceding results indicated that, while *ZNF518B* did not play any important role in cell proliferation, it did in cell migration and invasiveness. It has been described that migration of cancer cells through extracellular matrix most often occurs via type I collagen fibres. Therefore, the effects of the silencing of *ZNF518B* on the capacity of adhesion to type I collagen was next studied. It can be observed in Figure 70 that the adhesion capacity of CRC cells to type I collagen decreased in parallel with gene silencing (p-value <0.001). Therefore, it can be concluded that *ZNF518B* favours adhesion of CRC cells to collagen fibres.

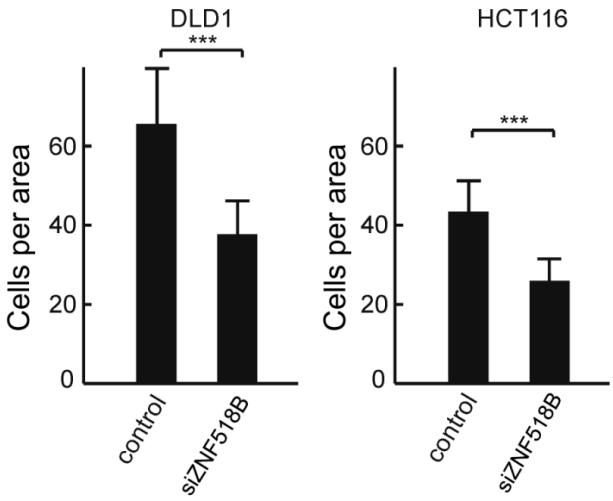

Figure 70: Adhesion test to type I collagen-coated plates. The effects of knocking down the gene *ZNF518B* on the adhesion of cells were quantified and the plots show the average of three cell counts. The results were compared by Student's t-test. ***, $p < 0.001$.

All of the above events, increase of cell motility, invasiveness and interaction with type I collagen fibres, are usually associated with epithelial-to-mesenchymal transition (EMT). Moreover, ZNF518B is a putative transcription factor, therefore, the effects of *ZNF518B* expression on the level of EMT markers were next examined. It can be observed in Figure 71 that the knocking down of *ZNF518B* caused a significant (p-value < 0.05) decrease in the expression of *SNAI1* in CRC cell lines and a significant (p-value <0.001) increase in the expression of *CDH1*. Also, it can be observed a significant (p-value <0.001) decrease in the N-cadherin and SNAIL protein level after *ZNF518B* silencing. This fact together with migration, invasion and adhesion functional assays suggests an implication of the *ZNF518B* gene in the development of EMT in CRC.

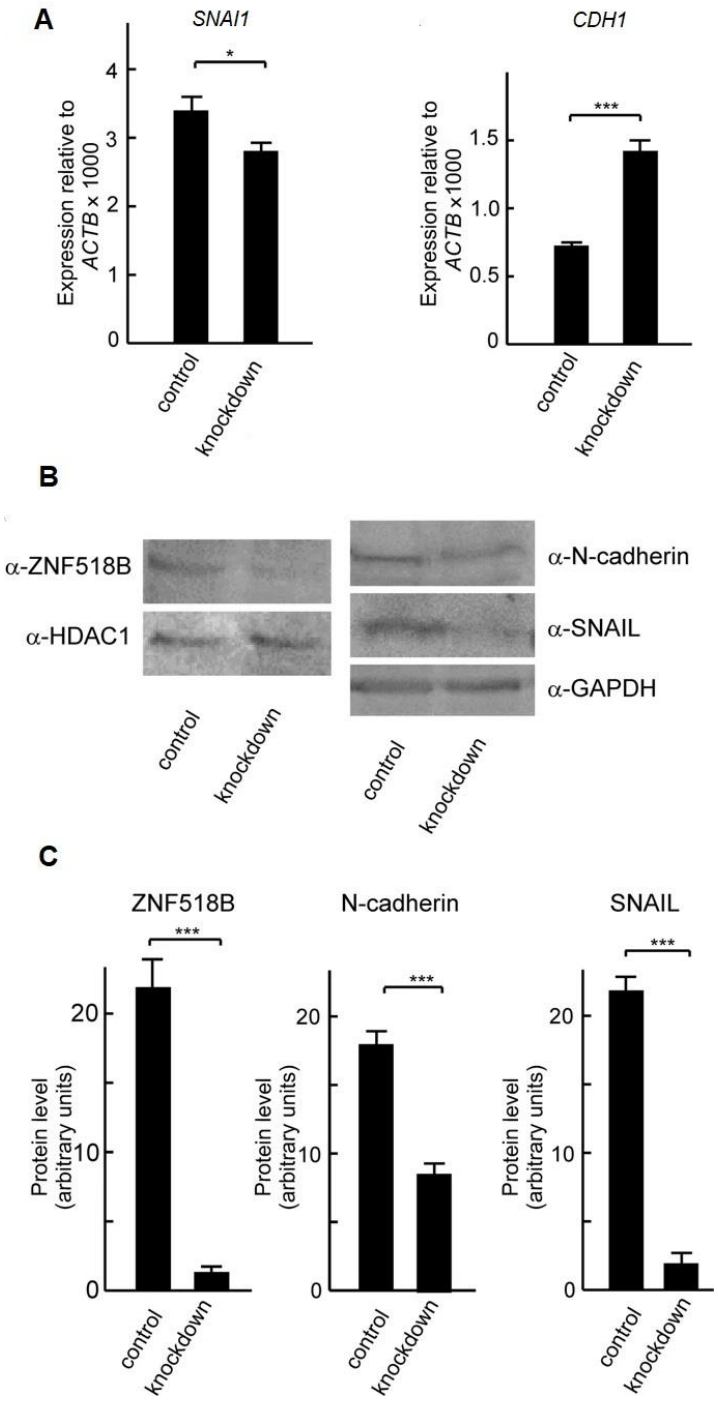

Figure 71: Effects of *ZNF518B* expression on the level of EMT markers. A) mRNA expression of *SNAI1* and *CDH1* in DLD1 cell line after silencing measure by RT-qPCR. B) Western blots showing the level of ZNF518B, N-cadherin and SNAIL after *ZNF518B* silencing. Histone deacetylase 1 (HDAC) or glyceraldehyde-3-phosphate dehydrogenase (GAPDH) were used as loading controls. Due to the different size of ZNF518B and of the EMT markers, the electrophoreses were carried out separately in 6% and 12% polyacrylamide gels, respectively. C) ImageJ analysis of the western blots of panel C. Four grey values relative to the loading controls were measured in every case and averaged. The results were compared by Student's t-test. *, $p < 0.05$; ***, $p < 0.001$.

4.2.2.4 Overexpression of *ZNF518B* supports the hypothesis of its involvement in epithelial-mesenchymal transition.

The knocking down experiments described in the previous paragraphs strongly supports the idea that *ZNF518B* is involved in the EMT. As a second strategy, to further explore whether this gene is related to EMT, an overexpression of *ZNF518B* was performed. The overexpression was achieved by transient transfection with a commercial plasmid, which contains the whole cDNA of the gene, named p*ZNF518B*. The information on the plasmid, as well as on the empty plasmid (pCAGGS) used as control, can be found under Material and Methods.

The following experiments were performed with CRC cell lines RKO in which, as mentioned, the *ZNF518B* is not expressed. To observe the increase in the expression of *ZNF518B* an immunocytochemistry experiment was first performed in RKO (Figure 72 A, B). It can be determined that the expression of *ZNF518B* is much more marked at the nuclear level in RKO transfected with p*ZNF518B* than in cells transfected with the empty plasmid. The overexpression of *ZNF518B* in nuclei from RKO was also detected in the western blots from nuclear extracts, which show the presence of a band of the size corresponding to the canonical isoform of ZNF518B in the cells transfected with p*ZNF518B* (Figure 72 C).

The increase in the expression of the gene after transfection with p*ZNF518B* was also checked at the mRNA level by RT-qPCR. This increase is clearly observed in RKO cells (Figure 73).

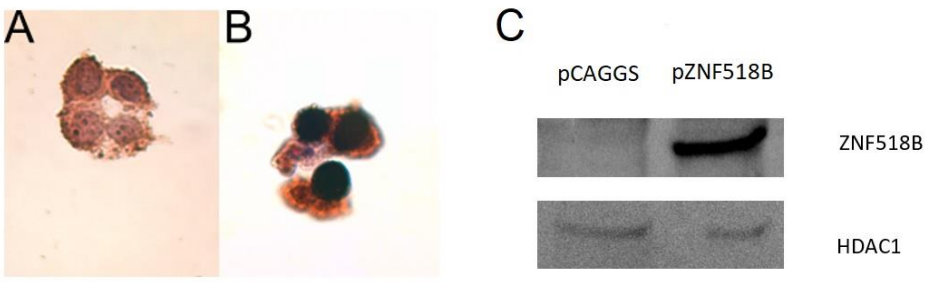

Figure 72: Detection by immunocytochemistry and western blot of the ZNF518B protein level in RKO cell line. A) RKO cells transfected with the empty plasmid (pCAGGS). B) RKO cells transfected with p*ZNF518B*. C) Detection of ZNF518B protein in nuclear extract from RKO cells by western blot. In the lower image, the nuclear housekeeping protein, HDAC1, is shown as loading control.

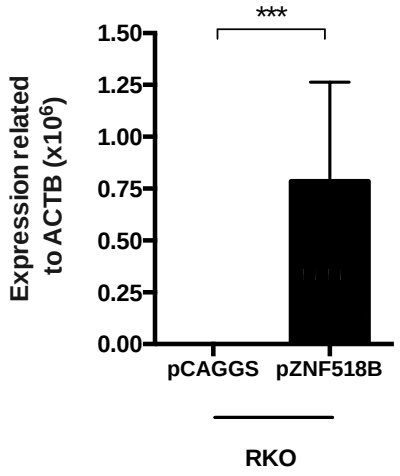

Figure 73: RT-qPCR showing the expression of *ZNF518B* in RKO cells transfected with the empty plasmid pCAGGS and with the plasmid p*ZNF518B*. The results were compared by Student's t-test. ***, $p < 0.001$.

Once the overexpression of ZNF518B was confirmed in RKO cells, the same functional assays done with silenced cells were performed with overexpressed cells to verify the role of this gene in EMT.

As regards the adhesion test, an increase in the adhesion of the RKO to type I collagen is observed after the increase of the expression of *ZNF518B* (Figure 74).

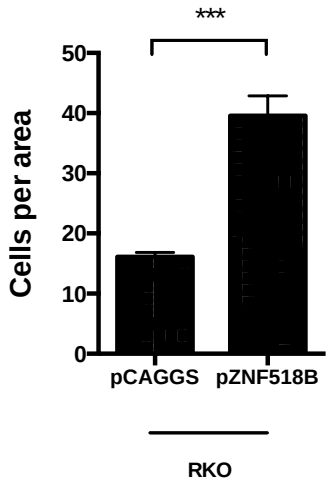

Figure 74: Adhesion test to type I collagen-coated plates after overexpression of ZNF518B in RKO cell line. The figure shows the comparison between RKO cells transfected with an empty plasmid (pCAGGS) and those transfected with the plasmid carrying ZNF518B. The results were compared by Student's t-test. ***, $p < 0.001$.

In parallel with the silencing experiments, the effects of *ZNF518B* overexpression on migration and invasion in RKO, was examined. Results are plotted and shown in Figure 75.

The effects of *ZNF518B* overexpression on colony formation, was next examined. The growth capacity of isolated colonies increases as the expression of *ZNF518B* in RKO was increased (Figure 76).

These results were not in agreement with the colony formation results obtained with silenced cells, in which no significant changes were observed (Figure 65). So, the role of *ZNF518B* gene in proliferation process is not clear and would need a more extensive study.

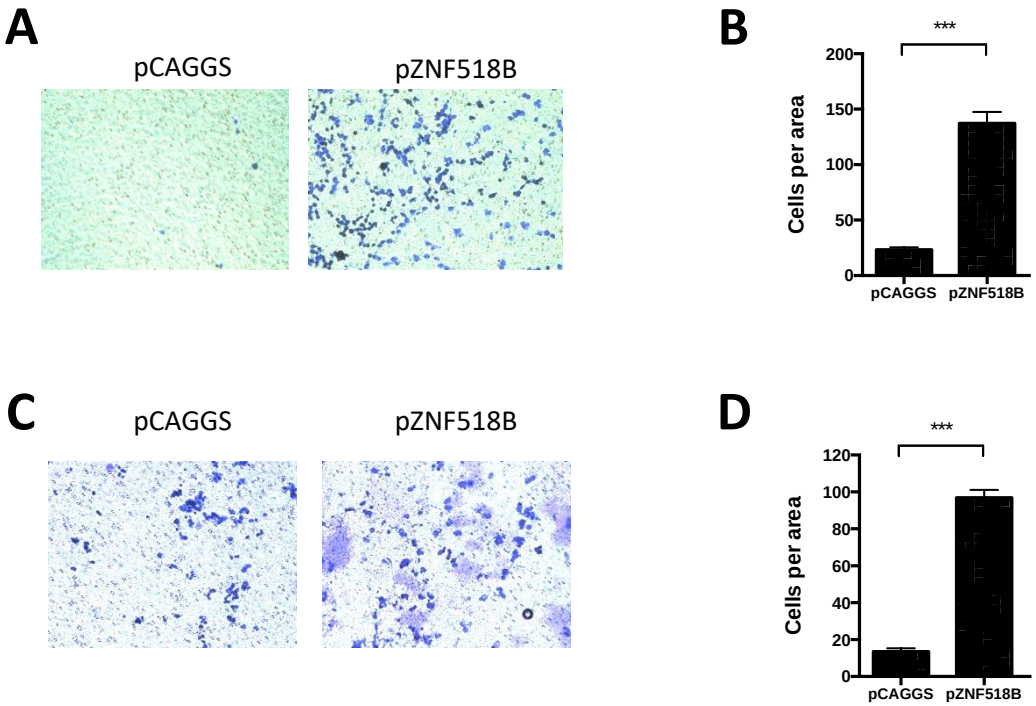

Figure 75: Transwell migration and invasion assay in RKO cell lines after overexpression of ZNF518B. Representative images (A) and quantitative analysis (B) showing the migration role of RKO cells due to overexpression of *ZNF518B* by the p*ZNF518B* in comparison with control (pCAGGS). Representative images (C) and quantitative analysis (D) showing the invasiveness role of RKO cells due to overexpression of *ZNF518B*. The quantitative analysis shows the averaged quantification of 3 fields in 3 plates each. The results were compared by Student's t-test. ***, $p < 0.001$.

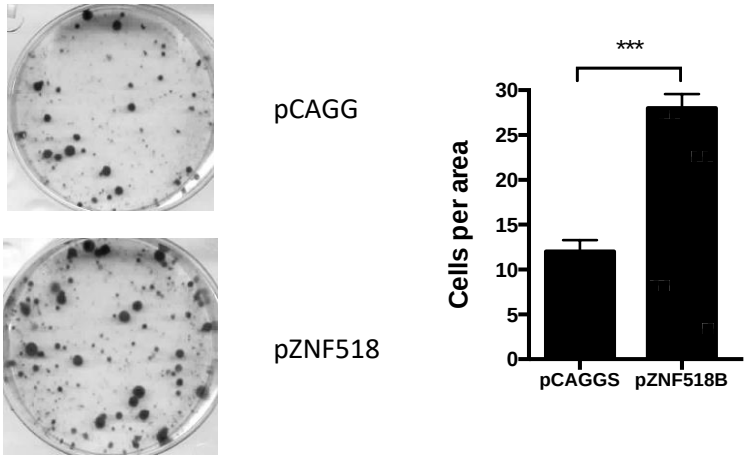

Figure 76: Colony formation test on RKO cell line after overexpression of ZNF518B. On the left, representative images and quantitative analysis showing the increase in colony forming capacity of RKO cells when transfected with a plasmid overexpressing *ZNF518B*. On the right, the quantitative analysis shows the averaged quantification of 3 fields in 3 plates each. The results were compared by Student's t-test. ***, $p < 0.001$.

4.2.2.5 *ZNF518B* as a regulator of gene expression in colorectal cancer cell lines.

ZNF518B is a zinc-finger protein, which presumably acts as a transcriptional factor. To conclude the functional study, it would be interesting to obtain some clues as to the genes putatively regulated by ZNF518B. To do this, a global transcriptomic analysis was carried out in DLD1 and HCT116 cell lines in which the *ZNF518B* gene was knocked down (see above). The assay, conducted by using a ClariomS (Affimetrix) analysis, revealed that the *ZNF518B* gene could be involved in the regulation of various genes related to different biological pathways.

Globally speaking, the silencing of *ZNF518B* caused the up-regulation of 163 and 276 genes in DLD1 and HCT116, respectively, as well as the down-regulation of 282 genes in DLD1 and 322 in HCT116. As shown in the Venn diagram of Figure 77, there are 64 genes commonly altered in both cell lines by the silencing of *ZNF518B* gene, corresponding to a 5.8% of the total of gene expression changes (Annex 1).

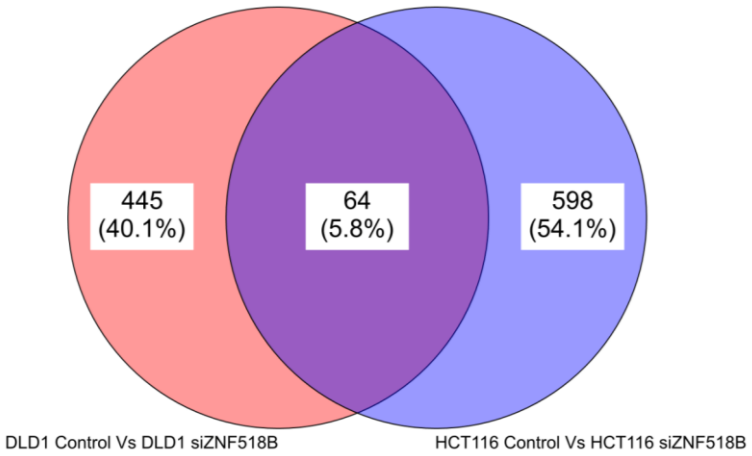

Figure 77: Venn diagram showing the number of genes, and the percentage, altered due the *ZNF518B* silencing. On the left the number of genes altered in DLD1 cell line is shown, while the results on the right corresponds to the HCT116 cell line. The intersection of both, in the middle of the diagram, shows the genes commonly altered in the DLD1 and HCT116 cell lines due the silencing of *ZNF518B* gene.

The Figure 78, depicts a global analysis of the data by principal components analysis (PCA). It is possible to observe the differences in the transcriptome between DLD1 and HCT116, and how the silencing of *ZNF518B* causes three-dimensional variations in the graph.

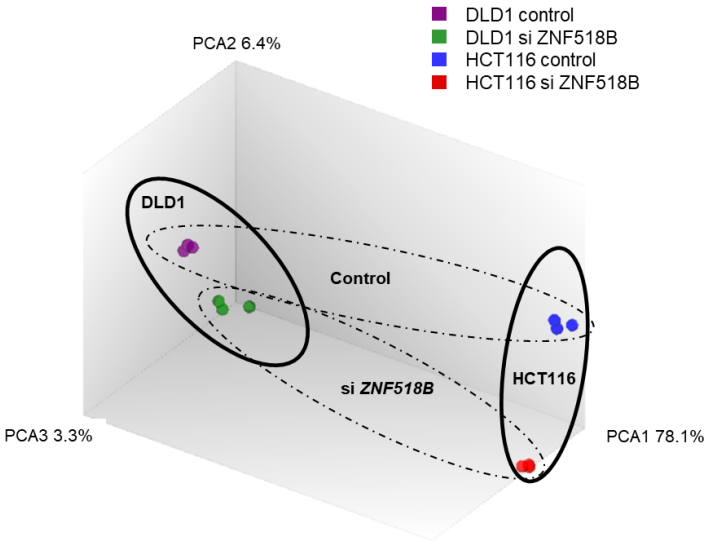

Figure 78: PCA diagram showing differences in the transcriptome profile of DLD1 and HCT116 when normal cells and cells in which *ZNF518B* was silenced, were compared.

It can be observed in the heatmap corresponding to the DLD1 (Figure 79), that the transcriptomic pattern is significantly different between the control and the gene-silencing conditions. These differences can be represented in the volcano plot of Figure 80, where a large cluster of significantly down-regulated genes (red dots) and up-regulated (green dots) can be observed.

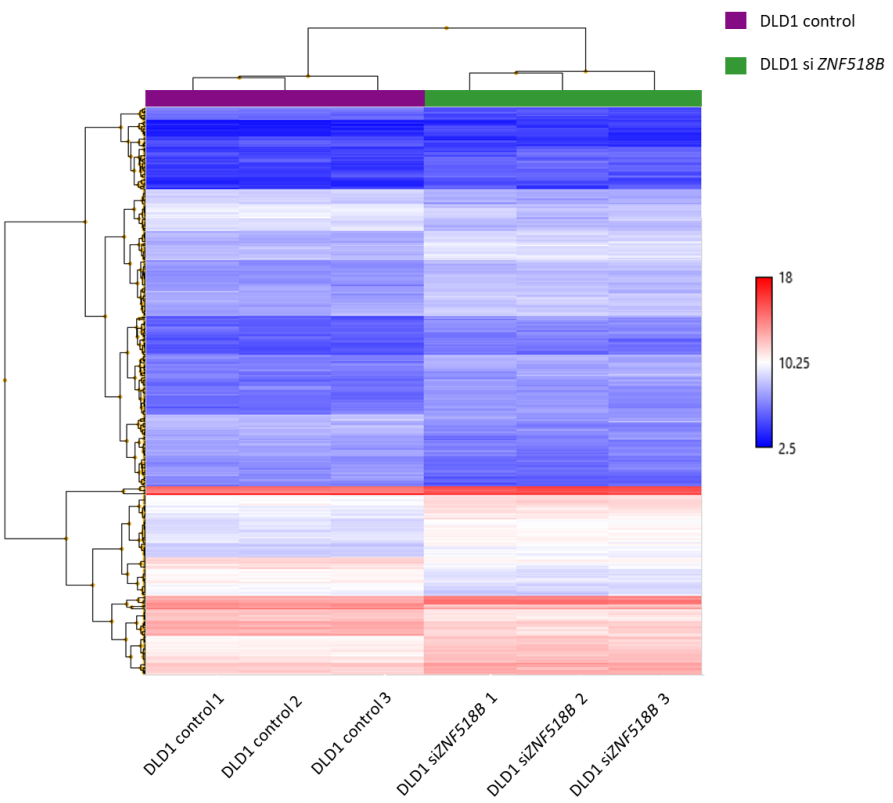

Figure 79: Heatmap of the transcriptomic variations in the cell line DLD1 caused by the silencing of *ZNF518B*. The triplicate of DLD1 control cell lines are shown on the left. The triplicate of DLD1 with *ZNF518B* knocked-down cell lines are shown on the right. Inside the heatmap, the red colour is related with the increase of gene expression; conversely, the blue colour is related with the decrease of gene expression.

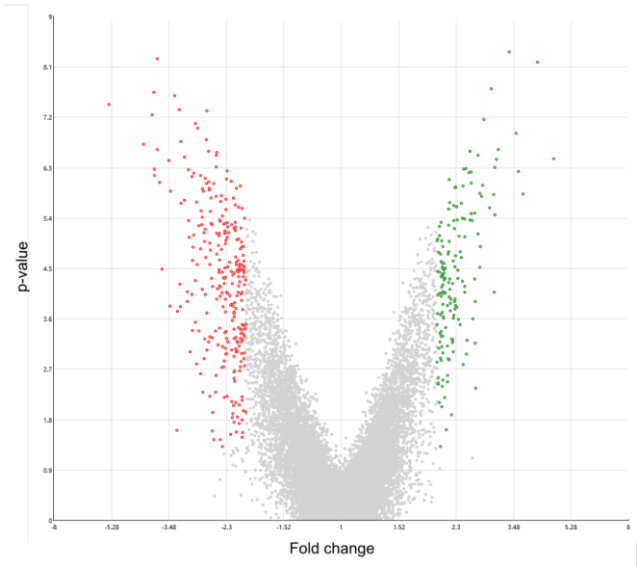

Figure 80: Volcano plot of the genes altered by silencing *ZNF518B* in the cell line DLD1. Down-regulated genes are shown as red dots and up-regulated genes as green dots.

Specifically, amongst the genes modified by silencing *ZNF518B* in DLD1 several of them were involved in various interesting biological pathways, like: the EMT process in CRC; the signalling in focal adhesion via PI3K-AKT-mTOR or the general focal adhesion, among others.

In the Table 11, some important genes and the biological processes in which they are involved were summarised. The rest of significantly altered genes after *ZNF518B* knocking-down in DLD1 were attached in the Annex 2.

Table 11: Selected genes altered by silencing of *ZNF518B* related to biological processes in DLD1 cell lines.

Biological processes	**Gene symbol**	**Regulation by *ZNF518B***
EMT in CRC	*TGFB2*	-
	NR2C2	-
	ITGA5	-
	JAG1	-
	TWIST1	+
	TWIST2	+
	FOXQ1	+
Focal adhesion	*LAMA3*	-
	THBS1-3	-
	ITAG5-9	-
	EPOR	-
	CREB5	-
	DDIT4	+
	SLC2A3	+
	RAC3	+
	DUSP1	+
	FOS	+
Cell cycle	*ATM*	-
	RAD21	-
	YWHAG	-
	CDC25C	-
	RBL2	+
Histones modification	*HIST1H3D*	+
	HIST1H4A	+
	HIST1H4C	+
Apoptosis	*HSPA1A*	+
	NAIP	-
Angiogenesis	*DDAH1*	-

The symbol "-" indicates a negative regulation by *ZNF518B*, conversely the symbol "+" indicates a positive regulation by *ZNF518B*.

Studies in HCT116 gave results similar to those in DLD1 knocking-down array, as can be seen in the corresponding heatmap (Figure 81). Many genes were significantly altered when the expression profile of control cells were compared with those of *ZNF518B*-silenced cells.

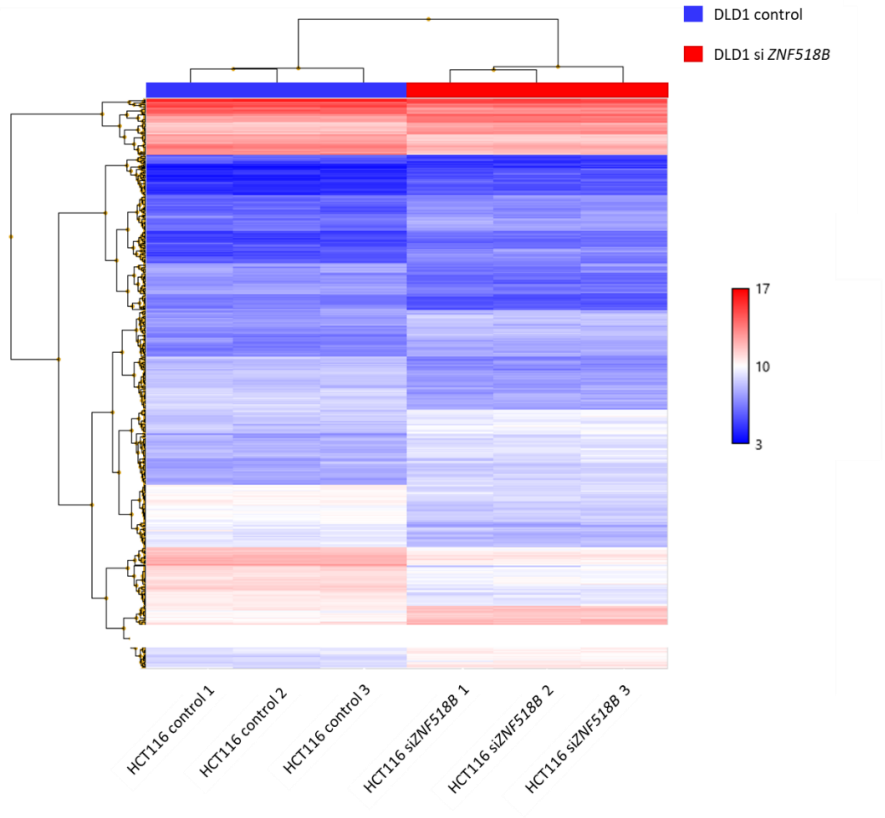

Figure 81: Heatmap of the transcriptomic variations in the cell line HCT116 caused by silencing of *ZNF518B*. The triplicate of HCT116 control cell lines are shown on the left. The triplicate of HCT116 with *ZNF518B* knocked-down cell lines are shown on the right. Inside the heatmap, the red colour is related with the increase of gene expression; conversely, the blue colour is related with the decrease of the gene expression.

Differences can also be represented as a volcano plot (Figure 82), in which a large cluster of significantly down-regulated genes (red dots) and up-regulated genes (green dots) can be observed.

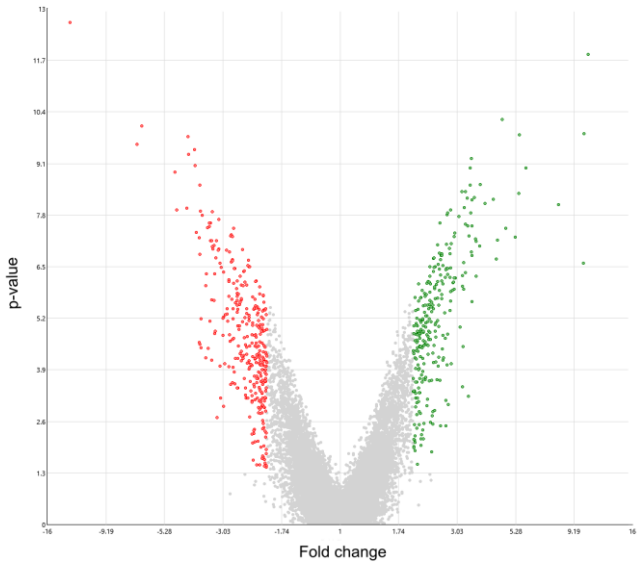

Figure 82: Volcano plot of the genes altered by the silencing of *ZNF518B* in the cell line HCT116. Down-regulated genes are shown as red dots and up-regulated genes as green dots.

The genes that have undergone variation in their expression due to the silencing of *ZNF518B* in HCT116 are numerous. Some selected genes related to important biological processes are observed in the Table 12. The rest of significantly altered genes due the *ZNF518B* knocking-down were attached in Annex 3.

Taking into account the function of *ZNF518B* in CRC regardless of the cell line, DLD1 or HCT116, a search should be carried out according only to two variables, the control and the cells in which *ZNF518B* has been silenced. In this case, in was found that a total of 2493 genes were significantly altered with a p-value less than 0.05, while 347 genes were significantly altered according to the FDR (False Discovery Rate) < 0.05 statistics. These data are shown in the heatmap of Figure 83.

Table 12: Selected genes altered by silencing of *ZNF518B* related to biological processes in HCT116 cell lines.

Biological processes	Gene symbol	Regulation by *ZNF518B*
EMT in CRC	*NOTCH1*	-
	TWIST1	-
	TWIST2	-
	SPARC	-
	ZEB1	-
Focal adhesion	*THBS1*	-
	EFNA5	-
	IGF2	-
	CREB5	-
Cell cycle	*CDK6*	+
	CDK7	+
	CCNE1	+
	RBL2	+
	PLK1	+
	YWHAH	+
	YWHAQ	+
Apoptosis	*CASP7*	-
	CASP2	+
	DFFA	+
	BCL2L1	+
	BCL11A	-
	IKBKG	+
Angiogenesis	*DDAH1*	-

The symbol "-" indicates a negative regulation by *ZNF518B*, conversely the symbol "+" indicates a positive regulation by *ZNF518B*.

It is important to highlight, at the level of biological pathways, that the alterations generated by the silencing of *ZNF518B*, are identical to those generated in the two cell lines separately. In addition, as can be seen in the heatmap of Figure 83, the automated clustering based on similarities of genetic profiles, has been done by separating the cell lines based on the genetic modification generated by silencing *ZNF518B* instead of the different cell lines used. All the important biological routes and processes are reflected in the gene ontology given in Table 13 and Figure 84.

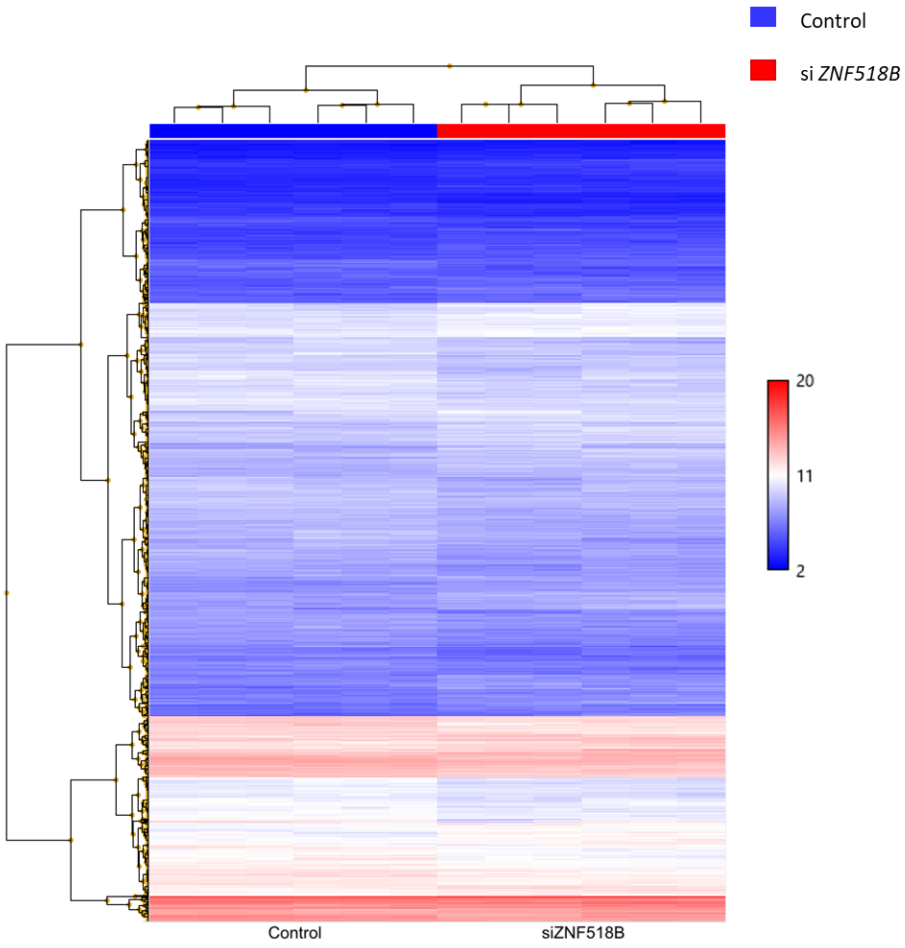

Figure 83: Heatmap showing the transcriptomic variations after silencing of *ZNF518B* in DLD1 and HCT116 cells. All the symbols as in figures 79 and 81.

Table 13: Gene ontology of the alterations caused by *ZNF518B* silencing in both cell lines.

Biological pathway	**Up-regulated genes (n)**	**Down-regulated genes (n)**
EGFR pathway	14	14
Signalling VEGFR2	16	27
PI3K-AKT pathway	26	17
MAPK pathway	21	18
WNT pathway	8	9
TGF-beta signalling pathway	13	13
RAS pathway	16	8
Biological processes	**Up-regulated genes (n)**	**Down-regulated genes (n)**
Cell cycle	9	20
EMT in CRC	12	9
Apoptosis	8	12
Histone modification	4	13
Focal adhesion	20	9
Focal adhesion via PI3K-AKT-mTOR	25	13

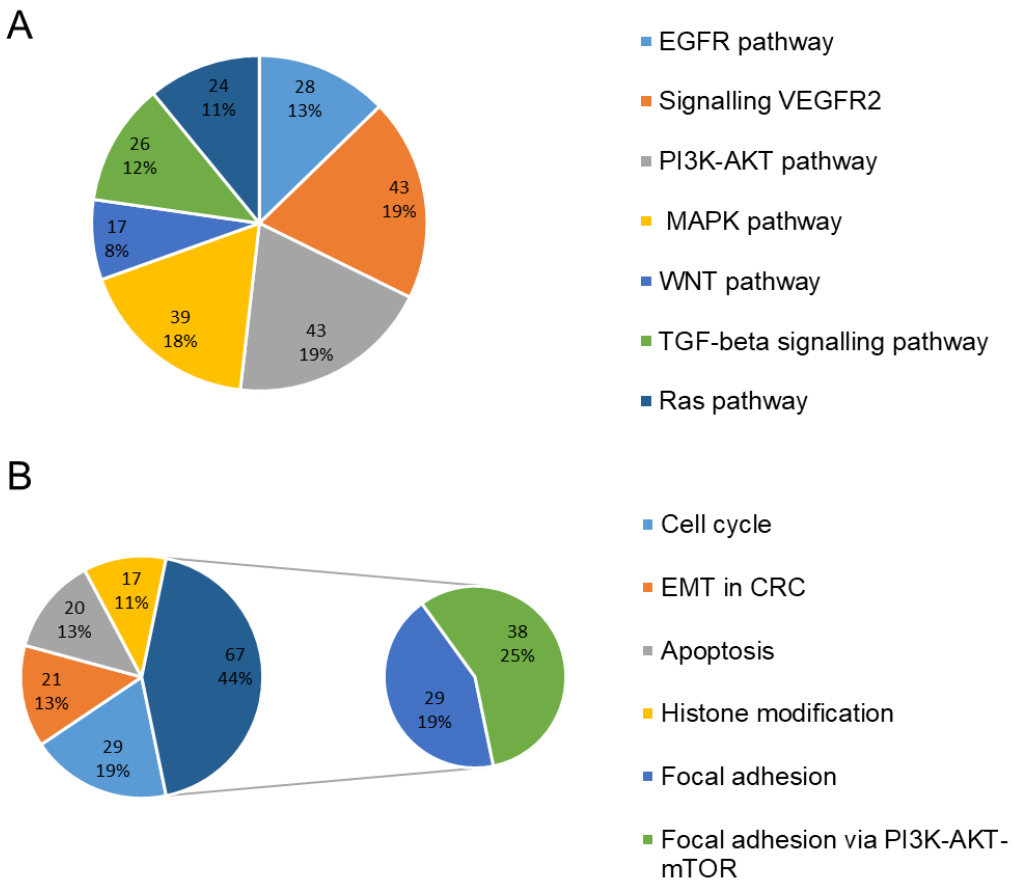

Figure 84: Pie chart of the alterations caused by *ZNF518B* silencing in both cell lines. A) Number and percentage of total genes altered in the main biological pathways. B) Number and percentage of total genes altered in the main biological processes.

To validate the results obtained by the array, RT-qPCR (Figure 85) of some selected genes was done. Selection was based on two criteria: first, those genes commonly altered in both cell lines and with greater fold change; and second, those genes that could be involved in progression of CRC based on the bibliography.

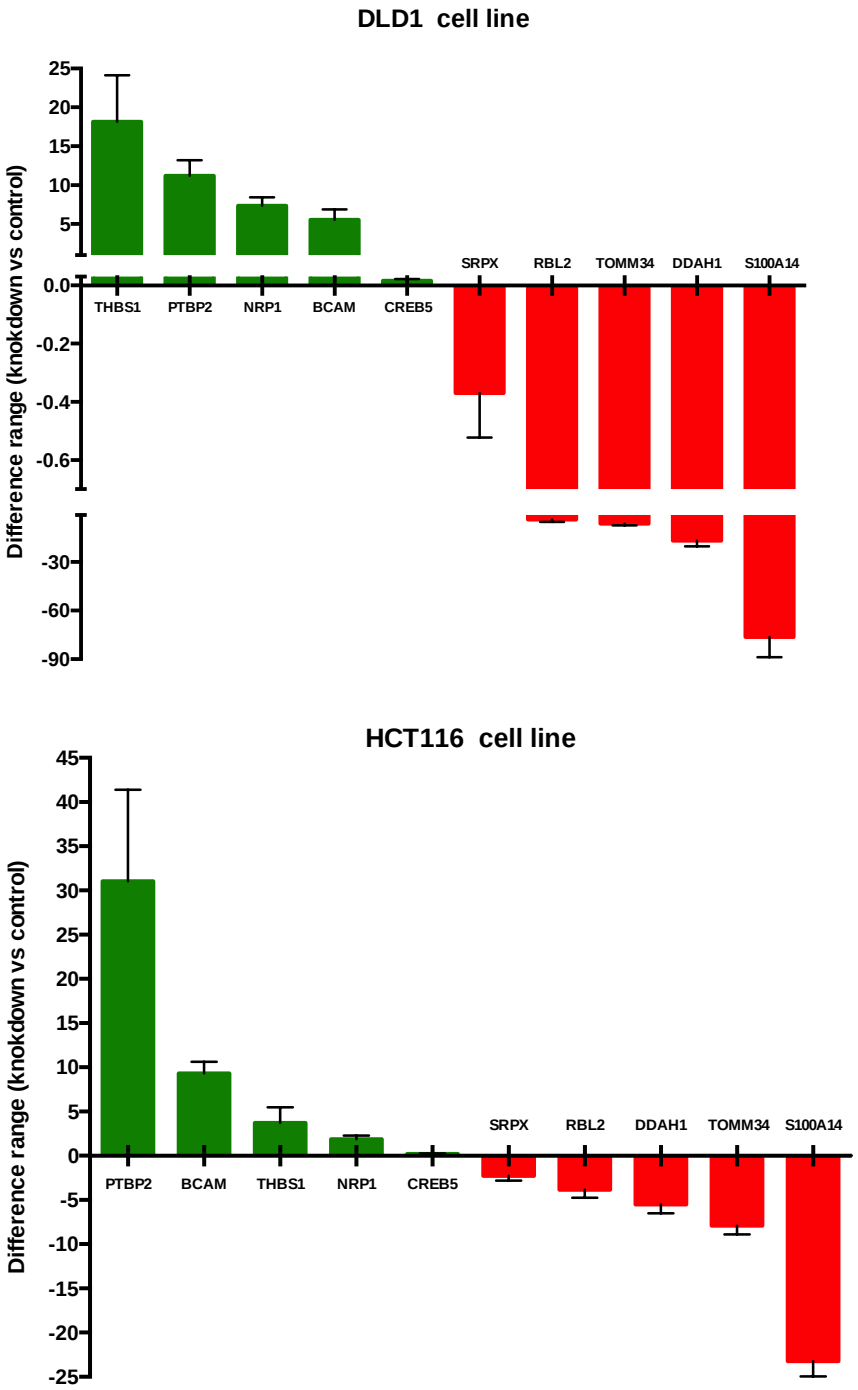

Figure 85: Validation by RT-qPCR of the genes detected in the array showing a greater fold change in expression in both cell lines. Note the scale cut on the ordinate axis in the DLD1 panel. The genes that are overexpressed after silencing *ZNF518B* are shown in green, while the genes under-expressed after silencing *ZNF518B* are marked in red.

4.2.3 Regulation of ZNF518B expression by epigenetic and chromatin structure features

As in the case of *EPDR1* gene, expression and functional analysis of *ZNF518B* gene suggest a role of this gene in progression of CRC. For this reason, it seemed interesting to explore the molecular causes involved in the regulation of *ZNF518B* expression. In this case, the methylation of the CpG island at the promoter could not be quantified, as commented under the Discussion. In view of this circumstance, other epigenetic mechanisms of regulation were studied, namely the chromatin structure and histone modifications. In both aspects our research group has a large experience. On the other hand, the influence of chromatin structure on cancer is an interesting and yet not common subject of research.

These studies may be adequately done through a comparison of the behaviour of SW48 and DLD1, because the gene is not transcribed in the former cell line, while it is in the second one. The D-Mut1 cell line, isogenic to DLD1, except that the wild type allele of *KRAS* has been knocked off, was also included in these experiments as a control, because the *ZNF518B* is also expressed in the latter cells.

As nucleosome organisation may play a regulatory role in gene expression, the chromatin structure at the promoter and proximal transcribed region was first studied. The results of a micrococcal nuclease protection assay from -1,000 to +300 relative to the transcription start site (TSS) are given in Figure 86 A, and the location of the tiled amplicons used can be seen in Figure 86 B.

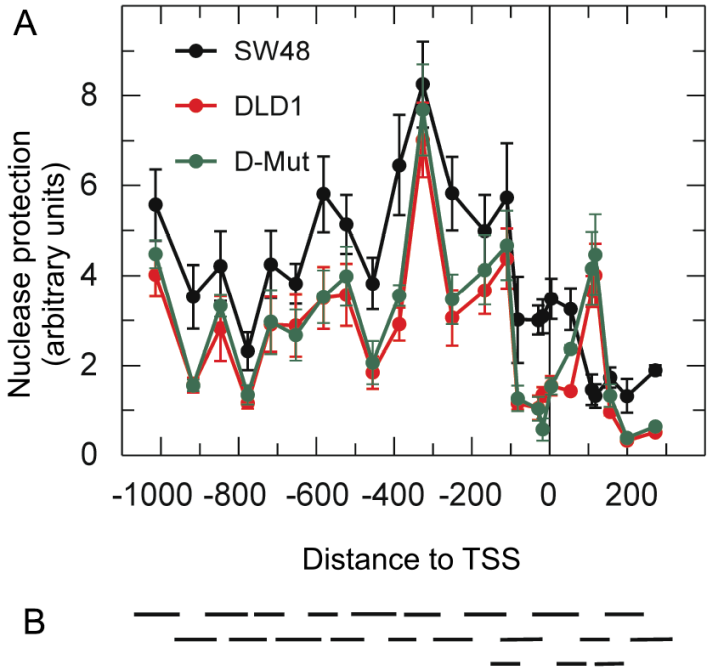

Figure 86: Organisation of chromatin at the promoter and proximal transcribed region of *ZNF518B* gene. A) Micrococcal nuclease protection assay. The protection, determined as described in the text, is plotted against the position of the centre of the amplicons used. The abscissae are given in bp relative to TSS. B) Location of the tiled amplicons.

Most of the peaks in Figure 86 A can be ascribed to the presence of nucleosomes as their width is compatible with nucleosomal size. A model for the nucleosomal organisation of the region under study is depicted in Figure 87. The nuclease-protected region between -500 and -800 may be ascribed either to two nucleosomes closely packed together or to a single nucleosome, N-3, which may occupy a series of sites due to sequence-directed, rotational positioning determinants. At any rate, as this aspect of chromatin structure is similar in the three cell lines, choosing one possibility or another is immaterial to the question of transcriptional regulation of the gene.

When examining the entire region analysed (Figure 86) it can be observed that the chromatin organisation in the three cell lines is very similar save in two details.

First, the overall protection to micrococcal nuclease is higher in the SW48 cell line, in which *ZNF518B* is not expressed. Second, there is a clear difference in the chromatin structure of the cell lines around the TSS. While in DLD1 and D-Mut1 cells the location of five nucleosomes (N+1 and N-4 to N-1) may be postulated and the TSS encompasses a nucleosome-free region, in SW48 cells the TSS is protected, suggesting that an additional nucleosome is positioned over it. Statistical analysis show that there is a significant difference in nuclease protection between SW48 and DLD1/D-Mut1 cells both in the region of TSS and around -200.

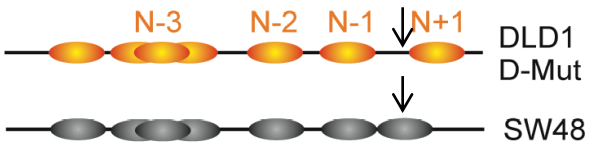

Figure 87: Nucleosomal localization model in the area surrounding TSS, based on the results obtained by the micrococcal nuclease protection assay. The negative nucleosomes (N-) correspond to the nucleosomes located 3' to the TSS point (black arrows), while the positive nucleosomes (N+) lie after it.

These results allow concluding that the structure of chromatin is important for the correct expression of the *ZNF518B* gene in CRC cell lines.

Epigenetic modification of histones is commonly associated with transcriptional regulation. Therefore, the presence of some epigenetic marks in the three cell lines was next studied. The determination was carried out by Nuc-ChIP, which allows the localisation of histone modifications at single nucleosome level. Acetylation at H3K9 and H3K27, associated with transcriptional activity, is characteristic of promoter nucleosomes. H3K4me3 and H3K9me3, which are usually associated, respectively, with active and repressed chromatin, are preferentially found in the proximal

transcribed region. Therefore, Nuc-ChIP to quantify H3K9ac and H3K27ac was done at amplicons -521, -324 and -108, which cover, respectively, nucleosomes N-3, N-2 and N-1; H3K4me3 and H3K9me3 were analysed at amplicon +75, which cover the first part of nucleosome N+1 in DLD1 and D-Mut1 cells and the second part of the nucleosome occupying the TSS in SW48. Figure 88 shows that nucleosomes N-3, N-2 and N-1 are enriched in H3K9ac and H3K27ac in DLD1 and D-Mut1 cells, while the marks are less abundant in SW48 cells. H3K4me3 is more abundant in nucleosome N+1 from the cells in which the gene is transcribed. On the contrary the mark H3K9me3 is especially present in the nucleosome covering the TSS in SW48 cells.

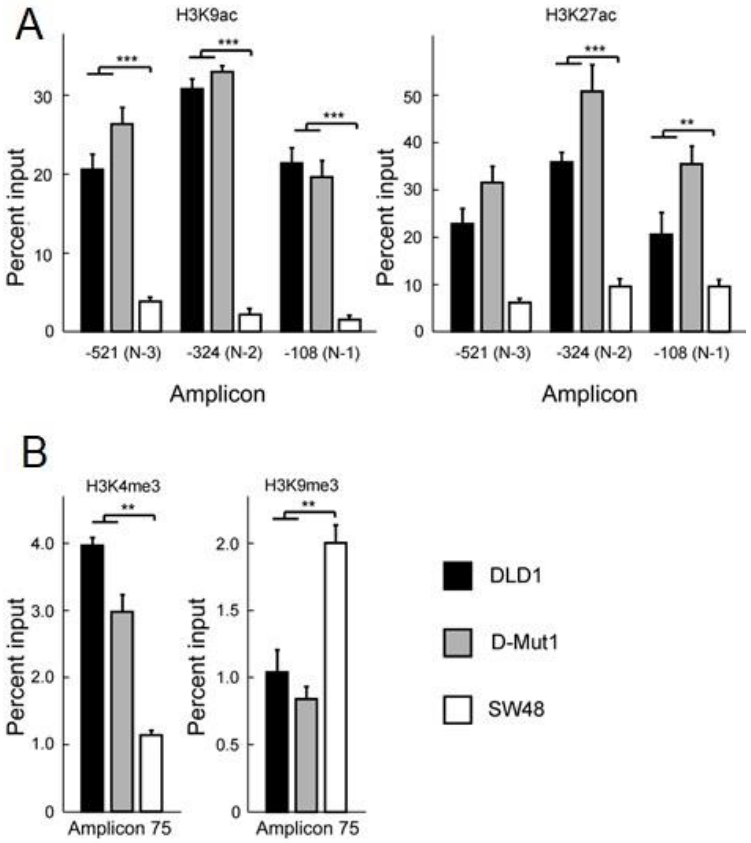

Figure 88: Epigenetic modification of histones in the promoter and proximal coding region of the *ZNF518B* gene. The experiments were carried out at single-nucleosome level by using the Nuc-ChIP approach. A) Acetylation marks in nucleosomes N-3, N-2 and N-1. B) Methylation marks at amplicon 75, which covers the 5' part of nucleosome N+1 in DLD1 and D-Mut1 cells and the 3' part of the nucleosome covering the TSS in SW48 cells. Results were compared with the Kruskal-Wallis test. (**. $p < 0.005$; ***, $p < 0.001$).

DISCUSSION

Discussion

Molecular subtyping based on gene expression is gaining interest in cancer stratification, and proposals for CRC have been done[3]. It has been recently emphasized that transcriptional signatures may allow a refinement of the CRC subtyping and, in this sense, finding novel genes dysregulated in the disease, either in tumour itself or in its environment, may aid to construct gene panels useful for diagnosis, prognosis or targeted therapy selection[24].

In a previous work, carried out in our laboratory, two genes, namely *EPDR1* and *ZNF518B*, were found to be expressed and its mRNA spliced in a manner dependent on the presence of mutations in *KRAS*. In the present work the expression of these genes has been studied in CRC patients and the mechanisms of their action and regulation have been investigated in an attempt to elucidate whether these genes may be included in the above referred panels. It is also worth studying the possible differential role of the various isoforms.

As seen in Figure 18 in the results section, performed by means of an array of CRC cDNA samples, the expression of *EPDR1* is clearly higher in tumour than in normal tissue. Also, the expression of isoform 2 of the gene was evaluated. This isoform has special characteristics that make it interesting with respect to the other isoforms of the gene. Based on the results obtained by Riffo *et al*. 2016 [158], the expression of isoform 2 is dependent on the mutation in *KRAS* G13D. The *KRAS* mutations in CRC are one of singular importance to know the malignancy of the tumour[38], as well as its behaviour in the administration of antitumour drugs [72] .

Isoform 2, unlike the other gene isoforms, is expressed to a sufficient level to allow being studied in patient samples. For its part, isoform 1, the major one, follows the behaviour, at the level of expression, of total *EPDR1*. Since isoform 2 is the isoform dependent on the mutation in *KRAS* in CRC cell lines, and this mutation is a factor of poor prognosis, we considered it appropriate to continue studying this isoform in the cohort of HCV patients.

As can be seen in the results, the bulk of the data obtained was made with the amplification of whole *EPDR1* gene, which includes isoform 1 and the interesting isoform 2, since the primers were designed within exons 4 and 5 of the gene. The paired samples, with which the bulk of the study was made, were paraffin-embedded, so that the quality of the sample prevented us from amplifying the isoform 2.

Based on the observations with paired samples, it can be concluded that, actually, the *EPDR1* gene is expressed more in tumour than in normal tissues. This observation corroborates that described in the bibliography, although the published studies were carried out only with three patient samples, being impossible to determine its significance[162].

Although the sample size is low, the immunohistochemistry results corroborate those obtained at mRNA level, in that in all cases the level of EPDR1 protein is higher in tumour that in non-tumour paired tissues.

It is relevant to observe that the higher expression of the *EPDR1* gene is detected in the advanced stages of the disease, that is, in the IV or metastatic stage. By multiple regression analysis clear evidence was obtained indicating that there is a significant relationship between the expression of *EPDR1* in tumour with respect to

healthy paired tissue. This relationship was observed either by difference in expression or by expression ratio.

The results of gene expression in tumour and adjacent normal tissues are usually given as the fold change, *i.e.*, the ratio of expression in tumour and adjacent tissue. Though useful for comparative purposes, some flaws may exist in this procedure, especially when the expression in normal tissues is very low or even below the detection limit. In these cases, the absolute error of the determination may result in large errors in the ratio, abnormally high values in it, or even a senseless infinite value. To prevent these errors, in the present thesis, apart from the customary fold change values, we have also statistically analysed the results subtracting the expression in normal tissues from that in tumour samples when the studies were carried out with paired samples.

Present results support the hypothesis that *EPDR1* is involved in CRC progression and metastasis. Overexpression of *EPDR1* in patients is associated with T and M parameters and, less significantly, with N. This is clearly in accordance with the results obtained after knocking down the gene in CRC cell lines. The thorough statistical analysis of the results of the present work reinforce the conclusions reached as to the involvement of *EPDR1* in CRC.

It is worth noting that patients' survival depends more on the expression of *EPDR1* in non-tumour adjacent mucosa than in tumour itself. The adjacent non-tumour tissues are defined as normal from a histological point of view, but they may be abnormal when their molecular characteristics are taken into account. For instance, it has been shown that histologically normal tissues surrounding breast cancer tissues exhibit some gene expression properties reflecting those of the cancer subtypes,

probably causing in some instances the recurrence after conservative therapies. Consequently, the importance of studying biomarkers in adjacent tissues to predict the risk of recurrence has been emphasized[196]. Recently, Aran *et al.*[98] have conducted an exhaustive transcriptomic study in which compared the tumour, non-tumour adjacent tissues and healthy tissues obtained by post-mortem collection. In some instances, specific gene activation occurs in adjacent tissues. In our case, as *EPDR1* overexpression is compatible with a diminution of cell adhesion and in higher cell association with type I collagen fibres, increased expression in adjacent tissues may facilitate dissemination and the risk of metastasis.

Returning to the importance of isoforms in the form of expression of the *EPDR1* gene, in many of the previously studied patients of the HCV cohort fresh tumour samples were available in the INCLIVA Biobank. Thanks to these fresh tissues, it has been possible to determine the level of isoform 2 and the ratio with respect to isoform 1. In all cases isoform 1 is the major one compared to 2, but it was interesting to deepen in the consequences of a differential expression of both isoforms. To do this, the ratio between the isoforms has been studied and Figure 26 shows the difference of the ratio of isoform 1 and isoform 2 of *EPDR1* in relation to the different clinicopathological parameters.

The results are very interesting because, significantly, the higher the ratio between isoform 2 and isofom 1, the higher the number of T4 patients, as well as the number of patients in stages II and stages III. Unfortunately, it was not possible to obtain samples from metastatic patients.

It is important to highlight the relationship between disease-free survival of the patients and the isoform ratio. There is a significant increase in relapsed patients when

the level of isoform 2 over isoform 1 is altered. Following this idea, a decrease in survival is observed based on a *Cox* regression model in those patients with a high I2/I1 ratio above the median. There are few genes in which the isoforms are related in any way to the survival of a patient cohort. A concrete example of this effect in a gene related to cancer is the insulin receptor, in which it has been observed that variations in its ratio of isoforms is related to different carcinogenic processes[197]. So, the study of the level of isoforms in spliced genes as well as the knowledge of their functions in the cell seems very important.

Thanks to the cohort of HCV patients, it has been possible to relate the level of *EDPR1* expression with the different subgroups of the CMS. As can be seen in Figure 55, the patients with higher expression of *EPDR1* show a clear trend towards being classified within the CMS3 group, that is, the molecular subgroup characteristic of mutations in *KRAS*. Although the sample size is low, the result is very important, since it increases the value of *EPDR1* as a candidate gene for inclusion within panels, not only for diagnosis and future treatments, but also for molecular classification of the patients. These data confirm the hypothesis drawn from the studies in cell lines[158], because the patients with altered expression of isoform 2 of *EPDR1*, are those that have mainly mutations in *KRAS*.

By observing the mean of the data, it can be seen that there is an important group of patients with highly expressed *EPDR1*, which are located within subtype 4. This subtype is characteristic of patients with alterations in genes involved in EMT, to which *EDPR1* can be added, based on the results presented in this thesis[21].

On the other hand, the mean *EPDR1* level of the patients classified within the CMS1 is very low. CMS1 is a subgroup that is characterized by having

hypermethylated tumors. Should this hypermethylation affect the CpG island studied in this thesis, the low expression of the *EDPR1* gene would be explained. To continue with the molecular subtyping, there are few patients having a mutation in *BRAF* and high expression value of *EPDR1*. In most cases, samples with a *BRAF* mutation have been classified as "low expression of *EPDR1*". *BRAF* is a characteristic gene within the classification group as CMS1[21].

As to the *ZNF518B* gene, the results shown in the thesis, allow concluding that both, isoform 1 and isoform 2 of the gene are overexpressed in tumour tissue with respect to normal tissue.

These results have been observed in a cDNA array of 5 samples of healthy colon mucosa and 43 CRC samples. Likewise, the same results have been corroborated with the HCV cohort. As indicated in the text, the expression of *ZNF518B* is human tissue, as in cell lines[158] is quite low. The quality of the cDNA prepared by retrotranscription of RNA isolated from paraffin-embedded samples, together with the low expression level of the gene, did not allow amplifying all the samples from the cohort by RT-qPCR. Therefore, only the results obtained with the canonical isoform from non-paired 70 tumour and 69 non-tumour adjacent samples were presented. As mentioned above, these results significantly show that the gene is overexpressed in CRC samples, but it was not possible to carry out a deep analysis of the results, as it was done in the case of *EPDR1*. The individual isoforms could neither be quantified in the paraffin-embedded samples, but they could be evaluated by using fresh tissue samples.

ZNF518B has 6 isoforms, of which only 1 and 2 generate protein. The isoform 2 depends on the *KRAS* mutation in cell lines[158]. The mRNA of this isoform is truncated and putatively generates a protein with only 75 amino acids.

The ratio isoform 2/isoform 1 could have a prognostic implication, since a significant relationship with patient's relapse rate could be identified (Figure 59). It can be speculated that this effect should rely on the differential interaction of the distinct isoforms: while isoform 1 contains both the methyl-transferase and the zinc finger domains, isoform 2 still interacts with G9a but lacking the zinc finger domain it can't bind DNA[188]. As a consequence, there will be a reduction of G9a methyl-transferase activity. Indeed, an increased ratio could be detected in those patients who did not relapse.

Once determined that *EPDR1* and *ZNF518B* are overexpressed in CRC tissues, it can be wondered what their function, if any, is. This question was addressed by *in vitro* experiments with CRC cell lines in which the genes were knocked down with siRNAs and overexpressed by transfecting the cells with the appropriate plasmids. As shown under Results, *EPDR1* favours cell proliferation, migration, invasiveness and adhesion to type I collagen fibres. Taking into account that these fibres have been described as the "highways" for tumour cell migration[199], the results obtained *in vitro* provide a mechanistic explanation to the association of *EPDR1* overexpression and the parameters T and M observed in patients. Present results disagree with those of Chu *et al*. 2018 [200], who claimed that *EPDR1* do not affect cell proliferation. The results described here in both, silencing and overexpression experiments, clearly show that *EPDR1* is involved in cell proliferation (Figures 30 and 31), in close relation to the results of the cell cycle analysis (Figure 32).

Based on the homology arguments posed by Gregorio-King *et al*. 2002[163], it has been assumed that mammalian ependymins are membrane proteins. It can be assumed that isoform 1, which possesses a signal peptide, as predicted by SignalP 5.0[201] (Figure 89) may indeed be a membrane protein. An experimental prove of this idea is provided by the immunocytochemical analysis carried out after overexpression of *EPDR1* in RKO cells (Figure 41). Nevertheless, as commented under the Introduction, EPDR1 has been identified as a lysosomal protein in rat brain[164]. These authors found that rat brain EPDR1 migrates, after complete deglycosylation, as a single band of ~22 kDa. The recent results of Wei *et al*. 2019 [165] also support the idea of a lysosomal localisation for EPDR1. To reconcile both types of evidence, it may be taken into account that the sequence of EPDR1 isoform 1 also possesses a couple of closely-spaced potential target sites for proteolytic removal of the signal peptide (Figure 89). Should this cleaving occur at the most probable of the two potential sites, the remaining peptide (38-224 in the EPDR1 isoform 1 sequence) would have a molecular weight of ~21.000, roughly similar to that of the lysosomal species. Therefore, it can be hypothesized that the EPDR1 isoform 1 has a dual destination: it can be targeted either to lysosomes to remain inside them after removal of the signal peptide, or to plasma membrane, where it can be interact, through its ectodomain, with extracellular matrix.

Taking for granted the presence of EPDR1 isoform 1 in membranes, the results described in this thesis, working with CRC cell lines, strongly suggest that the ectodomain of the protein diminishes intercellular adhesion. It has been described that the related piscine ependymins play a role in cell adhesion202. This question had not yet been studied with mammalian ependymin-like proteins, although most of the

pathologies in which EPDR1 is dysregulated or mutated (see Introduction) involve in one or another way a disturbance in cell adhesion.

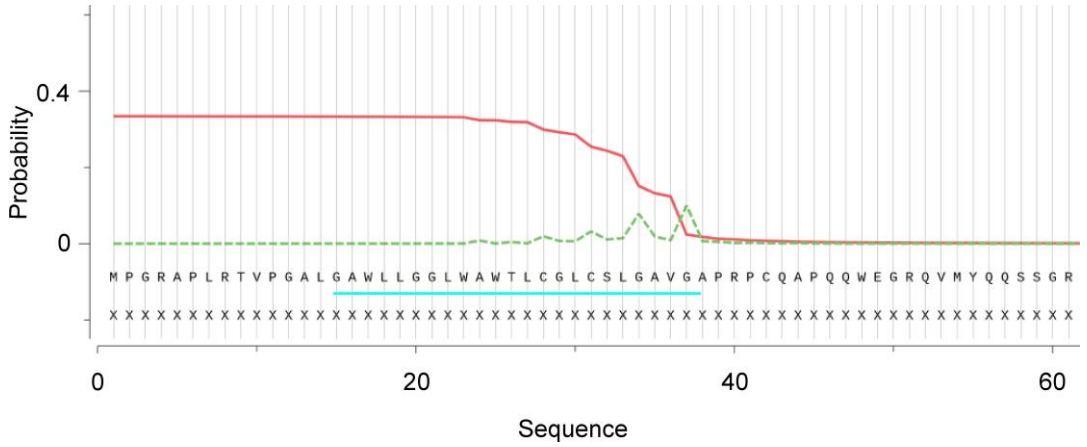

Figure 89: Prediction of EPDR1 segment function probability. The red line refers to the probability that a sequence corresponds to a membrane localization signal peptide. The green dashed line indicates the probability of the existence of a cut-off point for endoproteases. The blue line is postulated as a transmembrane segment based on sequence homology with ependymins[163]. Graph extracted from the SignalIP 5.0 predictor (Almagro, *et al*[201])

Obviously, EPDR1 is a protein involved in the development of cancer. It favours, as revealed by the T parameter, tumour malignancy promoting tumour growth and tissue infiltration, demonstrated by the *in vitro* experiments described here. EPDR1 was also studied based on the classification of the AJCC, the results are consistent with those obtained based on the TNM parameters, since we observed significant differences in stages II and IV, that is, in those in which the parameter N is not related. EPDR1, therefore, would not be related to the lymph node invasion.

Regarding to the functional studies of *ZNF518B* in CRC, first described in this thesis, are interesting, since the interaction of the canonical isoform with G9A could influence different processes of importance in the development of cancer. For

instance, the implications of the methyltransferase in the EMT have been reported[170] [171].

ZNF518B is not involved in cell growth or proliferation, unlike *EPDR1*, but a clear relationship with the promotion of migration and invasion, probably through the binding to type I collagen fibres is reported here. Results also agree with the role of G9A described, as it is related to invasiveness and metastasis in different types of cancer[189].

The results discussed up to now suggest that *ZNF518B*, as *EPDR1*, is involved in EMT. As this process is of crucial importance in cancer cell dissemination, it seemed interesting to check if this gene is actually involved in EMT. To do this, markers of the aforementioned process, such as SNAIL and E-cadherin, were determined. The cell lines that express *ZNF518B* have altered the transcription of *SNAI1*, with that of *CDH1* being repressed. As G9A represses the expression of *CDH1*, the gene encoding E-cadherin[171], the model presented in Figure 90, may explain how the dysregulation of *ZNF518B* could affect EMT through changes in the expression of *CDH1*.

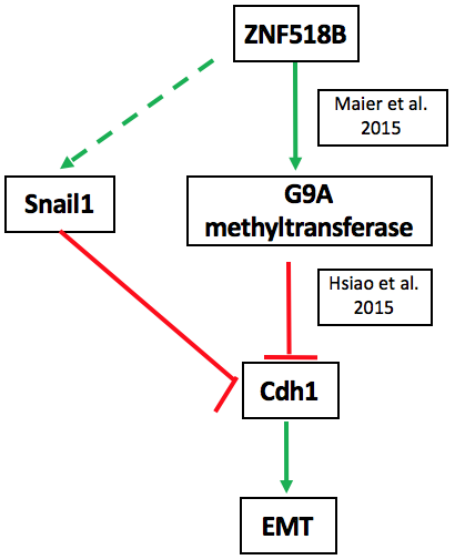

Figure 90: Model of a possible pathway linking overexpression of *ZNF518B* to EMT[188] [204].

ZNF518B is a putative transcriptional factor, and as such it must affect the activation or repression of a set of genes that could be involved in different routes or tumourigenic families. In this context, to explore in full the possible function of the gene in CRC, the possible targets of ZNF518B were analysed.

On the whole, thanks to the transcriptome-wide gene-level expression profiling tool named Clariom S, it was possible to determine that silencing of *ZNF518B* changes the expression of 64 genes common in both DLD1 and HCT116. When a gene ontology annotation for biological process was performed it was seen (Table 13) that *ZNF518B* affects genes involved in focal adhesion, EMT, modification of histones (H3 variants) and apoptosis. Many of these processes involve mechanisms in agreement with those demonstrated *in vitro* as commented in the Results section.

In relation to the genes more affected by the silencing of *ZNF518B*, some interesting observations can be made. There are several genes, highly altered by the silencing of *ZNF518B*, involved in the development of metastasis, related to migration and invasion in cancer. For instance, the gene *DDAH1* has been shown to be involved in the growth and in the promotion of angiogenesis in different types of tumours, such as glioma[205] prostate and breast[206] [207]. On the other hand, another marker of metastasis is *CREB5*, which is observed in the array as one of the genes with the greatest change after silencing *ZNF518B*. Specifically, this gene has been described to favour of metastasis in CRC[208]. In the same line, the gene *S100A14*, which encodes a protein with calcium binding functions, is overexpressed in many types of tumours and favours metastasis, as is the case of CRC[209]. Nevertheless, this is a conflicting question, as the gene has been described as suppressor of metastasis in other types of cancer such as gastric cancer[210]. Finally, among the genes to be highlighted due to

its fold-change and related with the metastatic processes and migration is the *NRP1* gene. There are studies showing the activity related to lymphocyte infiltration in CRC favouring liver metastases[211].

On the other hand, we observe another set of genes whose functions are involved in adhesion. In the Results section, we have already shown the role of *ZNF518B* in the adhesion to type I collagen-coated plates, when the gene is overexpressed in RKO cells or silenced in DLD1 and HCT166 cell lines. Among the genes involved in adhesion observed in the array as affected by the silencing of *ZNF518B* is *BCAM*. It codes for a membrane protein, overexpressed in many types of tumours. *LAMA3*, coding for a laminin subunit, has been also altered in the array and it has to be emphasized that genes coding for other laminin subunits have been found to favour metastasis in CRC, especially when *KRAS* is mutated[212 213]. A significant change in the *THBS1* gene has also been observed. It codes for thrombospondin 1, a membrane glycoprotein highly related to adhesion, and also to migration and tumourigenesis in different types of cancer[214]. Specifically, in CRC the overexpression of *THBS1* has been associated with the increase in the probability of relapse[215].

The role of some other genes affected by silencing *ZNF518B* must be highlighted. For instance, the genes *SRPX* and *RBL2*. In relation with the first one there are many articles in which it is associated with the decrease of proliferation, migration and invasion in prostate cancer[216 217] and in oesophageal cancer[218]. Specifically, in CRC, *SRPX* is found hypomethylated and its expression is elevated. It also favours angiogenesis via VEGF-A, and is related to a poor differentiation[219]. On the other hand, *RBL2* is part of the RB superfamily, key players in cell cycle processes and apoptosis through AKT[220]. *RBL2* has been studied in many types of cancer,

especially in lung[220]. The gene that displays the greatest fold-change after silencing of *ZNF518B* is *TOMM34* gene which encoded for a translocase of the mitochondrial membrane. This gene overexpressed in CRC, bladder[221] and breast cancer[222], has been proposed as a new therapeutic target[223].

The above considerations emphasize the relevance that the overexpression of *ZNF518B* gene may possess in the development of CRC as a transcriptional factor. Obviously, more research is needed to check whether the *ZNF518B*-induced oncogenic transformation occurs through changes in the expression of some of these putative gene targets.

In the preceding paragraphs the possible roles of *EPDR1* and *ZNF518B* in the onset and/or progression of CRC have been discussed. These roles made interesting to study the mechanisms of regulation of both genes. Taking into account that epigenetic dysregulation is very often a cause of neoplastic transformation[128], the epigenetic factors involved in the regulation of *EPDR1* and *ZNF518B* expression were studied in this thesis.

Regarding the regulation of *EPDR1*, we have shown that *EPDR1* expression is regulated by DNA methylation at a CpG island encompassing the first translatable exon in the major isoform. Methylation-dependent regulation of *EPDR1*, using a region shorter than ours, but overlapping with it, has also been recently reported by Chu *et al* [200]. Luo *et al.* found that in most of CRC stage IV patients the methylation of CpGs located between the positions that we show in Figure 45 is very low, although there is a subgroup of patients in which these CpGs are hypermethylated[224]. The results presented in this thesis show that the highest expression values of *EPDR1* are found in some stage IV patients and, therefore, our findings are in agreement with an

important, though not necessarily unique, role of DNA methylation in the regulation of *EPDR1* expression. In fact, in two of the cell lines analysed, DLD1 and D-Mut1, although they present a very similar hypomethylated profile (Figure 46), the expression level of the gene is different in both cell lines[158]

In order to explore whether other forms of gene regulation, such as that generated by microRNAs, could be implicated in fine-tune regulation of EPDR1 a bioinformatic tool[168] were used. It was observed that there is a match between both microRNA-193a-5p/*EPDR1* sequences. As observed in the results section, we were able to correlate the inverse expression of *EPDR1* with the level of 193a-5p microRNA. A decrease in miR-193a-5p expression has been correlated with an increase in the metastatic processes in CRC[167]. These results agree with the negative relationship of miR-193a-5p and of *EPDR1* expression (Figures 49 and 52) as well as with the proposed role of *EPDR1* in metastasis.

To verify this hypothetically direct interaction, a mimic, that is, an overexpressing agent of miR-193a-5p was used. It might be guessed that this would generate a decrease in the expression of *EPDR1* gene, in DLD1 and HCT116, but, as observed in Figure 50, this is not the case.

To explain this apparent contradiction, it should be reminded that, under normal physiological conditions the concentration of miR-193a-5p is the precise one to favour the inhibition/activation of its target genes. But when an exaggerate increase in the expression of the mimic occurs, the off-target effects and the possible anomalous unions induced by the abnormally high concentration can lead to aberrant effects. That is why, to avoid the action of off-target effects and to know how miR-193a-5p affects the expression of *EPDR1*, an over-expression was performed by *EPDR1* plasmids in

cell lines that do not express the gene (RKO cell lines). In this way, the artificial expression of EPDR1 is determined solely by the plasmid promoter. After checking the overexpression of *EPDR1*, a treatment was performed on the same cells with the mimic, which, as seen in the Figure 51 significantly reduces its expression, so the hypothesis of the inhibition of *EPDR1* expression by miR-193a-5p can be accepted.

In the case of *ZNF518B* gene, the structure of the CpG island in the promoter did not allow studying its methylation state, because a continuous array of CpG dinucleotides precludes the use of common methods to quantify methylation. The structure of chromatin, which is not studied frequently in cancer research, was instead determines, taking advantage of the large expertise of our laboratory in this kind of studies. The importance of this issue cannot be disregarded, because the regulation of eukaryotic gene expression has to be considered within the landscape of chromatin.

In this sense, the value of the experiments described here is reinforced, as they strongly suggest that the expression of *ZNF518B* is regulated by chromatin structure and epigenetic factors. The presence of a sequence-directed, positioned nucleosome covering the TSS in SW48 cells precludes the assembly of the pre-initiation complex resulting in transcriptional repression. On the contrary, a nucleosome-free region exists in *ZNF518B*-expressing cells around the TSS. Probably, the action of a chromatin-remodelling complex is responsible for these changes in chromatin structure and alterations in these remodelling machines have been described in many types of cancer[172]. Moreover, mutations in the ATPase subunit of SWI/SNF complex have been considered as drivers of cancer[173].

Apart from chromatin organisation, epigenetic modification of histones is also involved in the regulation of *ZNF518B* expression. The nucleosomes N-3, N-2 and N-

1 of DLD1 and D-Mut1 cells, in which the gene is transcribed, are enriched in H3K9ac and H3K27ac, a signature of promoter nucleosomes in active genes[174], whereas the level of these marks is significantly lower in SW48 cells.

The histone methylation marks at N+1 nucleosome are also compatible with their role in the regulation of gene expression. When these circumstances are considered together with the assumption that ZNF518B is involved in EMT, which, in turn, has been associated to modifications in the histone epigenetic marks[228], the *ZNF518B* gene might be added to the list of epigenetic targets for cancer therapy. Several therapeutic approaches are being developed in this sense [226 177] and it would be worth exploring these possibilities in the case of *ZNF518B* gene. Further work, based on much larger patients' cohorts, may decide its actual clinical value as a candidate for invasiveness prognosis in CRC, as well as its potentiality as a target for epigenetic drugs.

At any rate, present results may be useful to consider the inclusion of *EPDR1* and *ZNF518B* in panels of genes used to improve molecular subtyping of CRC. Also, *EPDR1* and *ZNF518B* expression or isoforms ratio of both as a bad prognostic marker could be an advance for improving the diagnosis of patients with CRC. On the other hand, it is important to highlight the role of the two genes in the increase of migration, invasion and adhesion cell capacity, therefore in the malignancy of the tumour cells. In addition, *EPDR1* is related to cell proliferation and both genes with the EMT. A deep knowledge of both genes will be required to fully exploit the therapeutic potential of targeting *EPDR1* and/or *ZNF518B* to treat colorectal cancer with special emphasis in the impact on metastasis and tissue invasion.

CONCLUSIONS

Conclusions

1.- The global expression of *EPDR1* and *ZNF518B* is significantly increased in tumor samples compared to normal samples in all the different cohorts of colorectal cancer patients analysed. Likewise, the alternative splicing form 2 of each of them increases in tumour tissue with respect to non-tumoural tissue in paired CRC samples.

2.- High-level expression of *EPDR1* is associated with the AJCC stages, T and M parameters and less significantly with N parameter in our HCV colorectal cancer patient cohort.

3.- An increase in the ratio between two splice isoforms of *EPDR1*, i2/i1, correlates significantly with T parameter, the AJCC stages and the *KRAS* mutational status. Particular attention should be paid to the fact that in both cases *EPDR1* and *ZNF518B* a significant alteration in the ratio of splice forms i2/i1 is observed in relapsing patients in our prospective cohort.

4.- Both, EPDR1 and ZNF518B have a role in migration, invasion and in adhesion capacity increasing the malignancy of the tumour cells *in vitro*. In addition, EPDR1 has also a role in increasing proliferation of cells though alteration of cell cycle and ZNF518B, for its part, regulates the expression of some EMT markers.

5.- The *EPDR1* expression is regulated by the CpG island methylation located in the first translatable exon of the gene. At the posttranscriptional level, microRNA 193a-5p, theoretically interacting with exon 6 of *EPDR1*, decrease

the *EPDR1* expression *in vitro*. Indeed, there is an inverse correlation between microRNA 193a-5p and *EPDR1* expression in HCV patient's cohort.
A relationship has been demonstrated between those patients who present a high degree of *EPDR1* expression with mutations in *KRAS* and *APC* and a low expression of *EPDR1* have been linked to the mutation in *BRAF*.

On the other hand, *ZNF518B* expression is regulated by chromatin structure of the promoter region and by different histone epigenetic modifications.

6.- High-level expression of *EPDR1* is clearly associated with two sugroups of the Consesus Molecular Classification of CRC, CMS3 and CMS4. Accordingly, with that, *KRAS* and *APC* mutated patients, characteristics from CMS3 subgroups, present higher EPDR1 expression levels. Also a low-level expression of *EPDR1* is found in patients classified as CMS1, fact that again agree with results showing a lower expression of *EPDR1* linked to the *BRAF* mutation, characteristic of CMS1 subgroup.

7.- A Microarray-based transcriptome analysis and gene ontology enrichment analysis reveals a role of *ZNF518B* in different biological process like cell cycle, EMT, apoptosis, histone modification and focal adhesion and in the protumourigenic pathways as EFGR, TGFβ, VEGFR2, WNT, PI3KCA, MAPK and KRAS, in two CRC cell lines models.

8.- Present results encourage the inclusion of *EPDR1* and *ZNF518B* in panels of genes used to improve molecular classification of CRC and also to further study both of them as a potential therapeutic targets to treat colorectal cancer with special emphasis in the impact on metastasis and tissue invasion.

REFERENCES

References

1. Fouad, Y. A. & Aanei, C. Revisiting the hallmarks of cancer. *Am J Cancer Res* **7**, 1016–1036 (2017).

2. Tariq, K., Ghias, K., Tariq, K. & Ghias, K. Colorectal cancer carcinogenesis: a review of mechanisms. *Cancer Biol Med* **13**, 120–35 (2016).

3. Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nat Med* **21**, 1350–1356 (2015).

4. Preet Kahai & Steve S. Bhimji. Anatomy, Abdomen and Pelvis, Large Intestine. *StatPearls Publ* (2018).

5. Andrew, M. & Hongmei, N. Epidemiology of colorectal cancer. *Int J Mol Epidemiol Genet* **7**, 105–114 (2016).

6. Arnold, M. *et al.* Global patterns and trends in colorectal cancer incidence and mortality. *Gut* **66**, 683–691 (2017).

7. Fleming, M., Ravula, S., Tatishchev, S. F. & Wang, H. L. Colorectal carcinoma: Pathologic aspects. *J Gastrointest Oncol* **3**, 153–173 (2012).

8. Van Cutsem, E., Cervantes, A., Nordlinger, B. & Arnold, D. Metastatic colorectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* **25**, iii1–iii9 (2014).

9. Haggar, F. A. & Boushey, R. P. Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. *Clin Colon Rectal Surg* **22**, 191–197 (2009).

10. Bhandari, A., Woodhouse, M. & Gupta, S. Colorectal cancer is a leading cause of cancer incidence and mortality among adults younger than 50 years in the USA: a SEER-based analysis with comparison to other young-onset cancers. *J Investig Med* **65**, 311–315 (2017).

11. MM, C., A, J., RA, S. & E, W. Worldwide variations in colorectal cancer. *CA Cancer J Clin* **59**, 366–378 (2009).

12. Rafiemanesh, H. *et al.* Epidemiology, incidence and mortality of lung cancer and their relationship with the development index in the world. *J Thorac Dis* **8**, 1094–1102 (2016).

13. JB, O. *et al.* Rates of colon and rectal cancers are increasing in young adults. *Am Surg* **69**, 866–872 (2003).

14. Ueno, H. *et al.* Optimal colorectal cancer staging criteria in TNM classification. *J Clin Oncol* **30**, 1519–1526 (2012).

15. Hari, D. M. *et al.* AJCC-7TH Edition Staging Criteria for Colon Cancer: Do the Complex Modifications Improve Prognostic Assessment? *J Am Coll Surg* **217**, 181–190 (2013).

16. Resch, A. & Langner, C. Lymph node staging in colorectal cancer : Old controversies and recent advances. *World J Gastroenterol* **19**, 8515–8526 (2013).

17. Kong, X. *et al.* A modified TNM staging system for non- metastatic colorectal cancer based on nomogram analysis of SEER database. *BMC Cancer* **50**, 10.1186/s12885-017-3796–1 (2018).

18. Lea, D., Haland, S., Hagland, H.R. & Soreider, K. Accuracy of TNM staging in colorectal cancer: a review of current culprits, the modern role of morphology and stepping-stones for improvements in the molecular era. *Scand J Gastroenterol* **49**, 1153–1163 (2014).

19. Li, J. *et al.* Staging of Colorectal Cancer Should be Reconsidered According to Weighting of the T Stage. *Medicine (Baltimore)* **95**, e2711 (2016).

20. Derwinger, K., Kodeda, K., Bexe-Lindskog, E. & Taflin, H. Tumour differentiation grade is associated with TNM staging and the risk of node metastasis in colorectal cancer. *Acta Oncol (Madr)* **49**, 57–62 (2010).

21. Arena, E. A., Bilchik, A. J., Wayne, J., Monica, S. & Angeles, L. What is the Optimal Means of Staging Colon Cancer? *Adv Surg* **47**, 199–211 (2015).

22. Mojarad, E. N., Kuppen, P. J. K., Aghdaei, H. A. & Zali, M. R. The CpG island methylator phenotype (CIMP) in colorectal cancer. *Gastroenterol Hepatol FRom Bed to Bench* **6**, 120–128 (2013).

23. Roelands, J. *et al.* Immunogenomic Classification of Colorectal Cancer and Therapeutic Implications. *Int J Mol Sci* **18**, 10.3390/ijms18102229 (2017).

24. Müller, M. F., Ibrahim, A. E. K. & Arends, M. J. Molecular pathological classification of colorectal cancer. *Virchows Arch* **469**, 125–134 (2016).

25. Dienstmann, R. *et al.* Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nature Reviews Cancer* **17**, 79–92 (2017).

26. Dickinson, B. T., Kisiel, J., Ahlquist, D. A. & Grady, W. M. Molecular markers for

colorectal cancer screening. *Gut* **64**, 1485–1494 (2015).

27. Sehgal, R. *et al.* Lynch Syndrome: An updated review. *Genes (Basel)* **5**, 497–507 (2014).

28. Kuipers, E. J. *et al.* Colorectal cancer. *Nat. Rev Cancer* **1**, (2015).

29. Pabla, B., Bissonnette, M. & Konda, V. J. Colon cancer and the epidermal growth factor receptor: Current treatment paradigms, the importance of diet, and the role of chemoprevention. *Word J Clin Oncol* **6**, 133–142 (2015).

30. Markman, B., Ramos, J., Capdevilla, J. & Tabernero, J. EGFR and KRAS in colorectal cancer. *Adv Clin Chem* **51**, 71–199 (2010).

31. Greally, M., Kelly, C. M. & Cercek, A. HER2: An emerging target in colorectal cancer. *Curr Probl Cancer* **42**, 560–571 (2018).

32. Moasser, M. M. The oncogene HER2; Its signaling and transforming functions and its role in human cancer pathogenesis. *Oncogene* **26**, 6469–6487 (2011).

33. W Burgess, A. EGFR family: Structure physiology signalling and therapeutic targets. *Growth Factors* **26**, 10.1080/08977190802312844 (2008).

34. Siena, S. *et al.* Targeting the human epidermal growth factor receptor 2 (HER2) oncogene in colorectal cancer. *Ann Oncol* **29**, 1108–1119 (2018).

35. Martinelli, E., Morgillo, F., Troiani, T. & Ciardello, F. Cancer resistance to therapies against the EGFR-RAS-RAF pathway: The role of MEK. *Cancer Treat Rev* **53**, 61–69 (2017).

36. Cohen, R. B. Epidermal Growth Factor Receptor as a Therapeutic Target in

Colorectal Cancer. *Clin Colorectal Cancer* **2**, 246–251 (2003).

37. Morilla, I., Diez, D., Moya-Garcia, A. A., Lozano, J. & Ranea, J. A. G. Exploring the interactions of the RAS family in the human protein network and their potential implications in RAS-directed therapies. *Oncotarget* **7**, (2016).

38. Simanshu, D. K., Nissley, D. V & McCormick, F. RAS Proteins and Their Regulators in Human Disease. *Cell* **170**, 17–33 (2017).

39. Sanz-Garcia, E., Argiles, G., Elez, E. & Tabernero, J. BRAF mutant colorectal cancer: Prognosis, treatment, and new perspectives. *Ann Oncol* **28**, 2648–2657 (2017).

40. Fang, J. Y. & Richardson, B. C. The MAPK signalling pathways and colorectal cancer. *Lancet Oncol* **6**, 322–327 (2005).

41. Bazley, L. A. & Gullick, W. J. The epidermal growth factor receptor family. *Endocr Relat Cancer* s17–s27 (2005).

42. Janku, F., Yap, T. A. & Meric-Bernstam, F. Targeting the PI3K pathway in cancer: are we making headway? *Nat Rev* **15**, 273–291 (2018).

43. Ruiz-Saenz, A. *et al.* HER2 amplification in tumors activates PI3K/Akt signaling independent of HER3. *Cancer Res* **78**, 3645–3658 (2018).

44. Danielsen, S. A. *et al.* Portrait of the PI3K/AKT pathway in colorectal cancer. *Biochim. Biophys Acta - Rev Cancer* **1855**, 104–121 (2015).

45. Molinari, F. & Frattini, M. Functions and Regulation of the PTEN Gene in Colorectal Cancer. *Front Oncol* **3**, 1–8 (2014).

46. Eklöf, V. *et al.* The prognostic role of KRAS, BRAF, PIK3CA and PTEN in colorectal cancer. *Br J Cancer* **108**, 2153–2163 (2013).

47. Francipane, M. G. & Lagasse, E. mTOR pathway in colorectal cancer: an update. *Oncotarget* **5**, 49–66 (2014).

48. Chang, F. *et al.* Involvement of PI3K/Akt pathway in cell cycle progression, apoptosis, and neoplastic transformation: A target for cancer chemotherapy. *Leukemia* **17**, 590–603 (2003).

49. Papadatos-Pastos, D., Rabbie, R., Ross, P. & Sarker, D. The role of the PI3K pathway in colorectal cancer. *Crit Rev Oncol Hematol* **94**, 18–30 (2015).

50. Polakis, P. Wnt signaling and cancer. *Genes Dev* **14**, 1837–1851 (2000).

51. Schneikert, J. & Behrens, J. The canonical Wnt signalling pathway and its APC partner in colon cancer development. *Gut* **56**, 417–425 (2007).

52. Fodde, R. The APC gene in colorectal cancer. *Eur J Cancer* **38**, 867–871 (2002).

53. Slattery, M. L. *et al.* Expression of Wnt-signaling pathway genes and their associations with miRNAs in colorectal cancer. *Oncotarget* **9**, 6075–6085 (2018).

54. Bendardaf, R., El-Serafi, A., Syrjänen, K., Collan, Y. & Pyrhönen, S. The effect of vascular endothelial growth factor-1 expression on survival of advanced colorectal cancer patients. *Libyan J Med* **12**, 1–5 (2017).

55. Nowak, D. G. *et al.* Regulation of Vascular Endothelial Growth Factor (VEGF) Splicing from Pro-angiogenic to Anti-angiogenic Isoforms. *J Biol Chem* **285**,

5532–5540 (2010).

56. Shibuya, M. VEGFR and type-V RTK activation and signaling. *Cold Spring Harb Perspect Biol* **5**, 1–13 (2013).

57. Koch, S. & Claesson-welsh, L. Signal Transduction by VEGFRs. *Cold Spring Harb Perspect Biol* **7** 10.1101/cshperspect.a00650 (2012).

58. Sánchez-Botet, A. *et al.* The atypical cyclin CNTD2 promotes colon cancer cell proliferation and migration. *Sci Rep* **8**, 1–12 (2018).

59. Simone, C., Resta, N., Bagella, L., Giordano, A. & Guanti, G. Cyclin E and chromosome instability in colorectal cancer cell lines. *J Clin Pathol Mol Pathol* **55**, 200–203 (2002).

60. Fang, Y. *et al.* Cyclin B1 suppresses colorectal cancer invasion and metastasis by regulating E-cadherin. *PLoS One* **10**, 1–16 (2015).

61. Helsten, T. *et al.* Cell-Cycle Gene Alterations in 4,864 Tumors Analyzed by Next-Generation Sequencing: Implications for Targeted Therapeutics. *Mol Cancer Ther* **15**, 1682–1690 (2016).

62. Sherr, C. J. & Bartek, J. Cell Cycle–Targeted Cancer Therapies. *Annu Rev Cancer Biol* **1**, 41–57 (2017).

63. Bendris, N., Lemmers, B. & Blanchard, J. M. Cell cycle, cytoskeleton dynamics and beyond: the many functions of cyclins and CDK inhibitors. *Cell Cycle* **14**, 1786–1798 (2015).

64. Harashima, H., Dissmeyer, N. & Schnittger, A. Cell cycle control across the

eukaryotic kingdom. *Trends Cell Biol* **23**, 345–356 (2013).

65. Zhou, N. & Gu, Q. Prognostic and clinicopathological value of p16 protein aberrant expression in colorectal cancer: A PRISMA-compliant Meta-analysis. *Med* **97**, e0195 (2018).

66. Yang, C. & Sun, J. J. Mechanistic studies of cyclin-dependent kinase inhibitor 3 (CDKN3) in colorectal cancer. *Asian Pacific J Cancer Prev* **16**, 965–970 (2015).

67. Barnum, K. J. & O'Connell, M. J. Cell Cycle Regulation by Checkpoints. *Methods Mol Biol* **1170**, 29–40 (2016).

68. Li, X. L., Zhou, J., Chen, Z. R. & Chng, W. J. P53 mutations in colorectal cancer-Molecular pathogenesis and pharmacological reactivation. *World J Gastroenterol* **21**, 84–93 (2015).

69. Armaghany, T., Wilson, J. D., Chu, Q. & Mills, G. Genetic Alterations in Colorectal Cancer. *Gastrointest Cancer Res* **5**, 19–27 (2012).

70. Nakayama, M. *et al.* Intestinal cancer progression by mutant p53 through the acquisition of invasiveness associated with complex glandular formation. *Oncogene* **36**, 5885–5896 (2017).

71. Muzny, D. M. *et al.* Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).

72. Martini, G. *et al.* Present and future of metastatic colorectal cancer treatment: A review of new candidate targets. *World J Gastroenterol* **23**, 4675–4688 (2017).

73. Healey, E. *et al.* Comparative effectiveness of 5-fluorouracil with and without

oxaliplatin in the treatment of colorectal cancer in clinical practice. *Anticancer Res* **33**, 1053–1060 (2013).

74. Longley, D. B., Harkin, D. P. & Johnston, P. G. 5-Fluorouracil: Mechanisms of action and clinical strategies. *Nat Rev Cancer* **3**, 330–338 (2003).

75. Calvo, E. *et al.* Irinotecan, oxaliplatin, and 5-fluorouracil/leucovorin combination chemotherapy in advanced colorectal carcinoma: A phase II study. *Clin Colorectal Cancer* **2**, 104–110 (2002).

76. Bao, X., Wu, J., Kim, S., LoRusso, P. & Li, J. Pharmacometabolomics Reveals Irinotecan Mechanism of Action in Cancer Patients. *J Clin Pharmacol* doi: 10.1002/jcph (2018).

77. Alcindor, T. & Beauger, N. Oxaliplatin: a review in the era of molecularly targeted therapy. *Curr Oncol* **18**, 18–25 (2011).

78. Liu, P., Cheng, H., Roberts, T. M. & Zhao, J. J. Targeting the phosphoinositide 3-kinase (PI3K) pathway in cancer. *Nat Rev Drug Discov* **8**, 627–644 (2009).

79. Jang, H. J., Kim, B. J., Kim, J. H. & Kim, H. S. The addition of bevacizumab in the first-line treatment for metastatic colorectal cancer: an updated meta-analysis of randomized trials. *Oncotarget* **8**, 73009–73016 (2017).

80. Hurwitz, H. I., Honeycutt, W., Haley, S. & Favaro, J. Long-term treatment with bevacizumab for patients with metastatic colorectal cancer: Case report. *Clin Colorectal Cancer* **6**, 66–69 (2006).

81. Scartozzi, M., Vincent, L., Chiron, M. & Cascinu, S. Aflibercept, a New Way to Target Angiogenesis in the Second Line Treatment of Metastatic Colorectal

Cancer (mCRC). *Target Oncol* **11**, 489–500 (2016).

82. Verdaguer, H., Tabernero, J. & Macarulla, T. Ramucirumab in metastatic colorectal cancer: evidence to date and place in therapy. *Ther Adv Med Oncol* **8**, 230–242 (2016).

83. Rogers, J. Patient considerations in metastatic colorectal cancer – role of panitumumab. *Onco Targets Ther* **10**, 2033–2044 (2017).

84. Van Cutsem, E. *et al.* ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. *Ann Oncol* **27**, 1386–1422 (2016).

85. Lee, M. S. *et al.* Efficacy of the combination of MEK and CDK4/6 inhibitors *in vitro* and *in vivo* in KRAS mutant colorectal cancer models. *Oncotarget* **7**, 39595–39608 (2016).

86. Cremolini, C. *et al.* FOLFOXIRI plus bevacizumab versus FOLFIRI plus bevacizumab as first-line treatment of patients with metastatic colorectal cancer: Updated overall survival and molecular subgroup analyses of the open-label, phase 3 TRIBE study. *Lancet Oncol* **16**, 1306–1315 (2015).

87. Cutsem, E. Van *et al.* Therapy for Patients With BRAF V600E – Mutant Metastatic Colorectal Cancer : Safety Lead-In Results From the Phase III BEACON Colorectal Cancer Study **37**, 22–24 (2019).

88. Sartore-Bianchi, A. *et al.* Dual-targeted therapy with trastuzumab and lapatinib in treatment-refractory, KRAS codon 12/13 wild-type, HER2-positive metastatic colorectal cancer (HERACLES): a proof-of-concept, multicentre, open-label, phase 2 trial. *Lancet Oncol* **17**, 738–746 (2016).

89. Lin, T., Leung, C., Nguyen, K. T. & Figlin, R. A. Mammalian target of rapamycin (mTOR) inhibitors in solid tumours. *Clin Pharm* **3**, 1–26 (2016).

90. Saura, C. *et al.* A first-in-human phase I study of the ATP-competitive AKT inhibitor Ipatasertib demonstrates Robust and safe targeting of AKT in patients with solid tumors. *Cancer Discov* **7**, 102–113 (2017).

91. Tolcher, A. W., Peng, W. & Calvo, E. Rational Approaches for Combination Therapy Strategies Targeting the MAP Kinase Pathway in Solid Tumors. *Mol Cancer Ther* **17**, 3–16 (2018).

92. Wang, M. *et al.* Role of tumor microenvironment in tumorigenesis. *J Cancer* **8**, 761–773 (2017).

93. Vinay, D. S. *et al.* Immune evasion in cancer: Mechanistic basis and therapeutic strategies. *Semin Cancer Biol* **35**, S185–S198 (2015).

94. Zamarron, B. F. & Chen, W. Dual Roles of Immune Cells and Their Factors in Cancer Development and Progression. *Int J od Biolocial Sci* **7**, 651–658 (2011).

95. Quail, D. & Joyce, J. Microenvironmental regulation of tumor progression and metastasis. *Nat Med* **19**, 1423–1437 (2013).

96. Belli, C. *et al.* Targeting the microenvironment in solid tumors. *Cancer Treat Rev* **65**, 22–32 (2018).

97. Gascard, P. & Tlsty, T. D. Carcinoma-associated fibroblasts: Orchestrating the composition of malignancy. *Genes Dev* **30**, 1002–1019 (2016).

98. Aran, D. *et al.* Comprehensive analysis of normal adjacent to tumor

transcriptomes. *Nat Commun* 10.1038/s41467-017-01027-z (2017).

99. Sanz-Pamplona, R. *et al.* Aberrant gene expression in mucosa adjacent to tumor reveals a molecular crosstalk in colon cancer. *Mol Cancer* **13**, 1–19 (2014).

100. Brabletz, T., Kalluri, R., Nieto, M. A. & Weinberg, R. A. EMT in cancer. *Nat Rev Cancer* **18**, 128–134 (2018).

101. Gout, S. & Huot, J. Role of cancer microenvironment in metastasis: Focus on colon cancer. *Cancer Microenviron* **1**, 69–83 (2008).

102. Pećina-Šlaus, N. Tumor suppressor gene E-cadherin and its role in normal and malignant cells. *Cancer Cell Int* **3**, 1–7 (2003).

103. Kwon, M. J. Emerging roles of claudins in human cancer. *Int J Mol Sci* **14**, 18148–18180 (2013).

104. Derycke, L. D. & Bracke, M. E. N-cadherin in the spotlight of cell-cell adhesion, differentiation, embryogenesis, invasion and signalling. *Int J Dev Biol* **48**, 463–476 (2004).

105. Liu, C.-Y., Lin, H.-H., Tang, M.-J. & Wang, Y.-K. Vimentin contributes to epithelial-mesenchymal transition cancer cell mechanics by mediating cytoskeletal organization and focal adhesion maturation. *Oncotarget* **6**, 15966–15983 (2015).

106. Niknami, Z., Eslamifar, A., Emamirazavi, A., Ebrahimi, A. & Shirkoohi, R. The association of vimentin and fibronectin gene expression with epithelial-mesenchymal transition and tumor malignancy in colorectal carcinoma. *Excli J* **16**, 1009–1017 (2017).

107. Ganesan, R., Mallets, E. & Gomez-Cambronero, J. The transcription factors Slug (SNAI2) and Snail (SNAI1) regulate phospholipase D (PLD) promoter in opposite ways towards cancer cell invasion. *Mol Oncol* **10**, 663–676 (2016).

108. Wang, Y., Shi, J., Chai, K., Ying, X. & Zhou, B. The Role of Snail in EMT and Tumorigenesis. *Curr Cancer Drug Targets* **13**, 963–972 (2013).

109. Choi, J. D. & Lee, J.-S. Interplay between Epigenetics and Genetics in Cancer. *Genomics Inform* **11**, 164 (2014).

110. Bardhan, K. & Liu, K. Epigenetics and colorectal cancer pathogenesis. *Cancers* **5**, 676–713 (2013).

111. Tse, J., Jenkins, L., Chionh, F. & Mariadason, J. Aberrant DNA Methylation in Colorectal Cancer: What Should We Target? *Trends Cancer* 10.1016/j.trecan.2017.08.003 (2017).

112. Zhang, W. & Xu, J. DNA methyltransferases and their roles in tumorigenesis. *Biomark Res* **5**, 1 (2017).

113. Esteller, M. CpG island hypermethylation and tumor suppressor genes: A booming present, a brighter future. *Oncogene* **21**, 5427–5440 (2002).

114. Brock, M. V., Herman, J. G. & Baylin, S. B. Cancer as a manifestation of aberrant chromatin structure. *Cancer J* **13**, 3–8 (2007).

115. Huang, T. *et al.* Targeting histone methylation for colorectal cancer. *Therap Adv Gastroenterol* **10**, 114–131 (2017).

116. Nakazawa, T. *et al.* Global histone modification of histone H3 in colorectal

cancer and its precursor lesions. *Hum Pathol* **43**, 834–842 (2012).

117. Loidl P. Towards an understanding of the biological function of histone acetylation. *FEBS Lett* 91–95 (1988).

118. Tse, J.W.T., Jenkins, L.J., Chionh, F. & Mariadason, J.M. Aberrant DNA Methylation in Colorectal Cancer: What Should We Target? *Trends Cancer* **3**, 698–712 (2017).

119. Noma, K., Allis, C.D., Grewall, S. Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. *Science* **293**, 1150–1155 (2001).

120. Khan, S. A. Global histone post-translational modifications and cancer: Biomarkers for diagnosis, prognosis and treatment? *World J Biol Chem* **6**, 333 (2015).

121. Simone, C. & Peserico, A. Physical and functional HAT/HDAC interplay regulates protein acetylation balance. *J Biomed Biotechnol* 10.1155/2011/371832 (2011).

122. Castillo, J., López-Rodas, G. & Franco, L. Histone Post-Translational Modifications and Nucleosome Organisation in Transcriptional Regulation: Some Open Questions. *Protein Rev* **18**, 62–95 (2017).

123. Vaiopoulos, A. G., Athanasoula, K. C. & Papavassiliou, A. G. Epigenetic modifications in colorectal cancer: Molecular insights and therapeutic challenges. *Biochim Biophys Acta - Mol Basis Dis* **1842**, 971–980 (2014).

124. Gezer, U. & Holdenrieder, S. Post-translational histone modifications in

circulating nucleosomes as new biomarkers in colorectal cancer. *In Vivo* **28**, 287–292 (2014).

125. Qin, J., Wen, B., Liang, Y., Yu, W. & Li, H. Histone Modifications and their Role in Colorectal Cancer (Review). *Pathol Oncol Res* 10.1007/s12253-019-00663-8 (2019).

126. Eberharter, A. & Becker, P. B. Histone acetylation: A switch between repressive and permissive chromatin. *EMBO Rep* **3**, 224–229 (2002).

127. Goossens-Beumer, I. J. *et al.* Age-dependent clinical prognostic value of histone modifications in colorectal cancer. *Transl Res* **165**, 578–588 (2015).

128. Rodríguez-Paredes, M. & Esteller, M. Cancer epigenetics reaches mainstream oncology. *Nat Med* **17**, 330–339 (2011).

129. Chi, Y. & Zhou, D. MicroRNAs in colorectal carcinoma - From pathogenesis to therapy. *J Exp Clin Cancer Res* **35**, 1–11 (2016).

130. Kenneth, K. T., Christy, W. T., Wu, M. & William, C. C. MicroRNAs in the prognosis and therapy of colorectal cancer: From bench to bedside. *Word J Gastroenterol* **24**, 2949–2973 (2018).

131. Tomasetti, M., Lee, W., Santarelli, L. & Neuzil, J. Exosome-derived microRNAs in cancer metabolism: Possible implications in cancer diagnostics and therapy. *Exp Mol Med* **49**, e285-11 (2017).

132. Svobodova, E., Kubikova, J. & Svoboda, P. Production of small RNAs by mammalian Dicer. *Pflugers Arch Eur J Physiol* **468**, 1089–1102 (2016).

133. Carthew, Richard W. and Sontheimer, E. J. Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136**, 642–655 (2010).

134. Yang, Y., Junjie, P., Sanjun, C. & Ma, Y. Long non-coding RNAs in colorectal cancer: Progression and future directions. *J Cancer* **8**, 3212–3225 (2017).

135. Poliseno, L. *et al.* A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **465**, 1033–1038 (2010).

136. Du, P. Q. *et al.* Down-regulation of ncRAN, a long non-coding RNA, contributes to colorectal cancer cell migration and invasion and predicts poor overall survival for colorectal cancer patients. *Mol Carcinog* **54**, 742–750 (2015).

137. Yosudjai, J., Wongkham, S., Jirawatnotai, S. & Kaewkong, W. Aberrant mRNA splicing generates oncogenic RNA isoforms and contributes to the development and progression of cholangiocarcinoma (Review). *Biomed Reports* **10**,147–155 (2019).

138. Zhu, L.Y., Zhu, Y.R., Dai, D.J., Wang, X. & Jin, H.C. Epigenetic regulation of alternative splicing. *Am J Cancer Res* **8**, 2346–2358 (2018).

139. Lee, Y. & Rio, D. C. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu Rev Biochem* **84**, 291–323 (2015).

140. Geuens, T., Bouhy, D. & Timmerman, V. The hnRNP family: insights into their role in health and disease. *Hum Genet* **135**, 851–867 (2016).

141. Cartegni, L., Chew, S. L., Krainer, A. R. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Gene* **3**, 285–298 (2002).

142. Will, C. L. & Lührmann, R. Spliceosome Structure and Function. *Cold Spring Harb Perspect Biol* **7**, a003707 (2011).

143. Huang, T. C. *et al.* Distinct roles and differential expression levels of Wnt5a mRNA isoforms in colorectal cancer cells. *PLoS One* **12**, 1–19 (2017).

144. Miura, K., Fujibuchi, W. & Unno, M. Splice isoforms as therapeutic targets for colorectal cancer. *Carcinogenesis* **33**, 2311–2319 (2012).

145. Pylayeva-Gupta, Y., Grabocka, E. & Bar-Sagi, D. RAS oncogenes: weaving a tumorigenic web. *Nat Rev Cancer* **11**, 761–774 (2011).

146. Riffo-Campos, Á. L. *et al.* Role of epigenetic factors in the selection of the alternative splicing isoforms of human KRAS in colorectal cancer cell lines. *Oncotarget* **9**, 20578–20589 (2018).

147. Dassano, A. *et al.* Mouse Pulmonary Adenoma Susceptibility 1 Locus Is an Expression QTL Modulating Kras-4A. *PLoS Genet* **10**, (2014).

148. Schubbert, S., Shannon, K. & Bollag, G. Hyperactive Ras in developmental disorders and cancer. *Nat. Rev Cancer* **7**, 295–308 (2007).

149. Schmick, M. *et al.* KRas localizes to the plasma membrane by spatial cycles of solubilization, trapping and vesicular transport. *Cell* **157**, 459–471 (2014).

150. Plowman, S. J. *et al.* The K-Ras 4A isoform promotes apoptosis but does not affect either lifespan or spontaneous tumor incidence in aging mice. *Exp Cell Res* **312**, 16–26 (2006).

151. Phipps, A. I. *et al.* KRAS-mutation status in relation to colorectal cancer survival:

The joint impact of correlated tumour markers. *Br J Cancer* **108**, 1757–1764 (2013).

152. Jinesh, G. G., Sambandam, V., Vijayaraghavan, S., Balaji, K. & Mukherjee, S. Molecular genetics and cellular events of K-Ras-driven tumorigenesis. *Oncogene* **37**, 839–846 (2018).

153. Jin, C. E. *et al.* Rapid and accurate detection of <i>KRAS</i> mutations in colorectal cancers using the isothermal-based optical sensor for companion diagnostics. *Oncotarget* **8**, 83860–83871 (2017).

154. Bai, B. *et al.* Mutations in KRAS codon 12 predict poor survival in chinese patients with metastatic colorectal cancer. *Oncol Lett* **15**, 3161–3166 (2018).

155. Fiala, O. *et al.* G12V and G12A KRAS mutations are associated with poor outcome in patients with metastatic colorectal cancer treated with bevacizumab. *Tumor Biol* **37**, 6823–6830 (2016).

156. Serra, R. W., Fang, M., Park, S. M., Hutchinson, L. & Green, M. R. A. KRAS-directed transcriptional silencing pathway that mediates the CpG island methylator phenotype. *Elife* e02313 (2014).

157. Roda, D. *et al.* EGF-Induced acetylation of heterogeneous nuclear ribonucleoproteins is dependent on KRAS mutational status in colorectal cancer cells. *PLoS One* **10**, 1–17 (2015).

158. Riffo-Campos, Á. L. *et al.* In silico RNA-seq and experimental analyses reveal the differential expression and splicing of EPDR1 and ZNF518B genes in relation to KRAS mutations in colorectal cancer cells. *Oncol Rep* **36**, 3627–3634

(2016).

159. E.Shashoua, V. Brain metabolism and the acquisition of new behaviors. III. Evidence for secretion of two proteins into the brain extracellular fluid after training. *Brain Res* **166**, 349–358 (1979).

160. Yin, H. & Flynn, A. D. Drugging Membrane Protein Interactions. *Annu Rev Biomed Eng* **18**, 51–76 (2016).

161. Staats, K. A., Wu, T., Gan, B. S., O'Gorman, D. B. & Ophoff, R. A. Dupuytren's disease susceptibility gene, EPDR1, is involved in myofibroblast contractility. *J Dermatol Sci* **83**, 131–137 (2016).

162. Nimmrich, I. *et al.* The novel ependymin related gene UCC1 is highly expressed in colorectal tumor cells. *Cancer Lett* **165**, 71–79 (2001).

163. Gregorio-King, C. C. *et al.* MERP1: A mammalian ependymin-related protein gene differentially expressed in hematopoietic cells. *Gene* **286**, 249–257 (2002).

164. Della Valle, M. C. *et al.* Demonstration of lysosomal localization for the mammalian ependymin-related protein using classical approaches combined with a novel density shift method. *J Biol Chem* **281**, 35436–35445 (2006).

165. Wei, Y. *et al.* Crystal structures of human lysosomal EPDR1 reveal homology with the superfamily of bacterial lipoprotein transporters. *Commun Biol* **2**, 1–13 (2019).

166. Shashoua, V. Ependymin, a brain extracellular glycoprotein, and CNS plasticity. *Ann N Y Acad Sci* **627**, 94–114 (1991).

167. Suárez-Castillo, E. C., Medina-Ortíz, W. E., Roig-López, J. L. & García-Arrarás, J. E. Ependymin, a gene involved in regeneration and neuroplasticity in vertebrates, is overexpressed during regeneration in the echinoderm Holothuria glaberrima. *Gene* **334**, 133–143 (2004).

168. Suga, T. *et al.* Influence of Multiple Genetic Polymorphisms on Genitourinary Morbidity After Carbon Ion Radiotherapy for Prostate Cancer. *Int J Radiat Oncol Biol Phys* **72**, 808–813 (2008).

169. Tilli, T. M., Bellahcène, A., Castronovo, V. & Gimba, E. R. P. Changes in the transcriptional profile in response to overexpression of the osteopontin-c splice isoform in ovarian (OvCar-3) and prostate (PC-3) cancer cell lines. *BMC Cancer* **14**, 1–11 (2014).

170. Behbehani, A. E., Ranjbari, N., Rahim, F. & Jazayeri, N. Evaluating the expression of CD34 marker in colorectal adenocarcinoma and its relationship with clinicopathologic factors. *Asian J Cell Biol* **10**, 80–86 (2015).

171. Malik, A., Mishra, R. N., Fanthome, B., Rao, R. & Patrikar, S. R. Role of CD34, vascular endothelial growth factor, and p53 in neoangiogenesis as correlated with stage of disease in colorectal carcinoma. *Med J Armed Forces India* **67**, 320–325 (2011).

172. Laity, J. H., Lee, B. M. & Wright, P. E. Zinc finger proteins new insights into structural and functional. *Curr Opin Struct Biol* **11**, 39–46 (2001).

173. Lu, D., Searles, M. A. & Klug, A. Crystal structure of a zinc-finger–RNA complex reveals two modes of molecular recognition. *Nature* **426**, 96–100 (2003).

174. Klug, A. The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation. *Annu Rev Biochem* **79**, 213–31 (2010).

175. Pavletich, N. P. & Pabo, C. O. Zinc Finger-DNA Recognition: Crystal structure of a Zif268-DNA Complex. *New Ser* **252**, 809–817 (1991).

176. Jantz, D. & Berg, J. M. Probing the DNA-binding affinity and specificity of designed zinc finger proteins. *Biophys J* **98**, 852–860 (2010).

177. Lentjes, M. H. *et al.* The emerging role of GATA transcription factors in development and disease. *Expert Rev Mol Med* **18**, 1–15 (2016).

178. Akiyama, Y. *et al.* GATA-4 and GATA-5 transcription factor genes and potential downstream antitumor target genes are epigenetically silenced in colorectal and gastric cancer *Mol Cell Biol* **23**, 8429–8439 (2003).

179. Katz, J. P. *et al.* The zinc-finger transcription factor Klf4 is required for terminal differentiation of goblet cells in the colon. *Development* **129**, 2619–2628 (2002).

180. Lee, H. young *et al.* High KLF4 level in normal tissue predicts poor survival in colorectal cancer patients. *World J Surg Oncol* **12**, 1–7 (2014).

181. Cassandri, M. *et al.* Zinc-finger proteins in health and disease. *Cell Death Discov* **3**, 17071 (2017).

182. Vu, T. & Datta, P. K. Regulation of EMT in colorectal cancer: A culprit in metastasis. *Cancers* **9**, 1–22 (2017).

183. Yang, L. *et al.* The previously undescribed ZKSCAN3 (ZNF306) is a novel

'driver' of colorectal cancer progression. *Cancer Res* **68**, 4321–4330 (2008).

184. Jen, J. & Wang, Y. C. Zinc finger proteins in cancer progression. *J Biomed Sci* **23**, 1–9 (2016).

185. Jin, T. bo *et al.* Genetic variations in the CLNK gene and ZNF518B gene are associated with gout in case–control sample sets. *Rheumatol Int* **35**, 1141–1147 (2015).

186. Zhang, X. Y. *et al.* SLC2A9 and ZNF518B polymorphisms correlate with gout-related metabolic indices in Chinese Tibetan populations. *Genet Mol Res* **14**, 9915–9921 (2015).

187. Bacos, K. *et al.* Blood-based biomarkers of age-associated epigenetic changes in human islets associate with insulin secretion and diabetes. *Nat Commun* **7**, 1–13 (2016).

188. Maier, V. K. *et al.* Functional Proteomic Analysis of Repressive Histone Methyltransferase Complexes Reveals ZNF518B as a G9A Regulator. *Mol. Cell. Proteomics* **14**, 1435–46 (2015).

189. Shankar, S. R. *et al.* G9a, a multipotent regulator of gene expression. *Epigenetics* **8**, 16–22 (2013).

190. Casciello, F., Windloch, K., Gannon, F. & Lee, J. S. Functional role of G9a histone methyltransferase in cancer. *Front Immunol* **6**, 3–9 (2015).

191. Wang, X., Spandidos, A., Wang, H. & Seed, B. PrimerBank: A PCR primer database for quantitative gene expression analysis, 2012 update. *Nucleic Acids Res* **40**, 1144–1149 (2012).

192. Riffo-Campos, Á. L. *et al.* Nucleosome-specific, time-dependent changes in histone modifications during activation of the early growth response 1 (Egr1) gene. *J Biol Chem* **290**, 197–208 (2015).

193. Kim, S. K. *et al.* A nineteen gene-based risk score classifier predicts prognosis of colorectal cancer patients. *Mol Oncol* **8**, 1653–1666 (2014).

194. Climente-González, H., Porta-Pardo, E., Godzik, A. & Eyras, E. The Functional Impact of Alternative Splicing in Cancer. *Cell Rep* **20**, 2215–2226 (2017).

195. Agarwal, V., Bell, G. W., Nam, J.W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* **4**, 1–38 (2015).

196. Casbas-Hernandez, P. *et al.* Tumor intrinsic subtype is reflected in cancer-adjacent tissue. *Cancer Epidemiol Biomarkers Prev* **24**, 406–414 (2015).

197. Vella, V., Milluzzo, A., Scalisi, N. M., Vigneri, P. & Sciacca, L. Insulin receptor isoforms in cancer. *Int J Mol Sci* **19**, 1–20 (2018).

198. Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nat Med* 10.1038/nm.3967 (2015).

199. Egeblad, M., Rasch, M. G. & Weaver, V. M. Dynamic interplay between the collagen scaffold and tumor evolution. *Curr Opin in Cell Biol* **22**, 697–706 (2010).

200. Chu, C.H. *et al.* Prognostic Values of EPDR1 Hypermethylation and Its Inhibitory Function on Tumor Invasion in Colorectal Cancer. *Cancers* **10**, 393 (2018).

201. Almagro Armenteros, J. J. *et al.* SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* **37**, 420–423 (2019).

202. Hoffmann, W. & Schwarz, H. Ependymins: Meningeal-Derived Extracellular Matrix Proteins at the Blood-Brain Barrier. *Int rev cyt* **165**, 121–158 (1996).

203. Liu, S. *et al.* G9a is essential for EMT-mediated metastasis and maintenance of cancer stem cell-like characters in head and neck squamous cell carcinoma. *Oncotarget* **6**, 6887–6901 (2015).

204. Hsiao, S. M. *et al.* The H3K9 Methyltransferase G9a Represses E-cadherin and is Associated with Myometrial Invasion in Endometrial Cancer. *Ann Surg Oncol* **22**, 1556–1565 (2015).

205. Kostourou, V., Robinson, S. P., Cartwright, J. E. & Whitley, G. S. J. Dimethylarginine dimethylaminohydrolase I enhances tumour growth and angiogenesis. *Br J Cancer* **87**, 673–680 (2002).

206. Reddy, K. R. K. *et al.* Novel Cellularly Active Inhibitor Regresses DDAH1 Induced Prostate Tumor Growth by Restraining Tumor Angiogenesis through Targeting DDAH1/ADMA/NOS Pathway. *ACS Comb Sci* **21**, 241–256 (2019).

207. Hulin, J. A., Tommasi, S., Elliot, D. & Mangoni, A. A. Small molecule inhibition of DDAH1 significantly attenuates triple negative breast cancer cell vasculogenic mimicry in vitro. *Biomed Pharmacother* **111**, 602–612 (2019).

208. Qi, L. & Ding, Y. Involvement of the CREB5 regulatory network in colorectal cancer metastasis.tle. *Yi Chuan* **36**, 679–84 (2014).

209. Wang, H.Y. *et al.* Expression status of S100A14 and S100A4 correlates with metastatic potential and clinical outcome in colorectal cancer after surgery. *Oncol Rep* **23**, 45–52 (2010).

210. Zhu, M. *et al.* Calcium-binding protein S100A14 induces differentiation and suppresses metastasis in gastric cancer. *Cell Death Dis* **8**, e2938 (2017).

211. Chaudhary, B. & Elkord, E. Novel expression of Neuropilin 1 on human tumor-infiltrating lymphocytes in colorectal cancer liver metastases. *Expert Opin Ther Targets* **19**, 147–61 (2015).

212. Bartolini, A. *et al.* BCAM and LAMA5 mediate the recognition between tumor cells and the endothelium in the metastatic spreading of KRAS-mutant colorectal cancer. *Clin Cancer Res* **22**, 4923–4933 (2016).

213. Chang, J. W. & Koo, B. S. LAMB3 mediates metastatic tumor behavior in papillary thyroid cancer by regulating c-MET/Akt signals. *Sci Rep* **8**,1068–1068 (2018). doi:10.1158/1538-7445.am2018-1068

214. Pal, S. K. *et al.* THBS1 is induced by TGFB1 in the cancer stroma and promotes invasion of oral squamous cell carcinoma. *J Oral Pathol Med* **45**, 730–739 (2016).

215. Waisberg, J. *et al.* Overexpression of the ITGAV gene is associated with progression and spread of colorectal cancer. *Anticancer Res* **34**, 5599–5607 (2014).

216. Hong, X., Hong, X., Zhao, H. & He, C. Knockdown of SRPX2 inhibits the proliferation, migration, and invasion of prostate cancer cells through the PI3K/Akt/mTOR signaling pathway. *J Biochem Mol Toxicol* **33**, 1–7 (2019).

217. Zhang, M. *et al.* High SRPX2 protein expression predicts unfavorable clinical outcome in patients with prostate cancer. *Onco Targets Ther* **11**, 3149–3157

(2018).

218. He, F. *et al.* SRPX2 knockdown inhibits cell proliferation and metastasis and promotes chemosensitivity in esophageal squamous cell carcinoma. *Biomed Pharmacother* **109**, 671–678 (2019).

219. Øster, B. *et al.* Non-CpG island promoter hypomethylation and miR-149 regulate the expression of SRPX2 in colorectal cancer. *Int J Cancer* **132**, 2303–2315 (2013).

220. Pentimalli, F. *et al.* RBL2/p130 is a direct AKT target and is required to induce apoptosis upon AKT inhibition in lung cancer and mesothelioma cell lines. *Oncogene* **37**, 3657–671 (2018).

221. Ahmed, M., Ali, M. H., Elatrash, G. A. & Abbas, H. H. TOMM34 expression and clinicopathological correlations in urothelial carcinoma of the urinary bladder. *Virchows Arch.* **469**, S237 (2016).

222. Aleskandarany, M. A. *et al.* TOMM34 expression in early invasive breast cancer: A biomarker associated with poor outcome. *Breast Cancer Res. Treat.* **136**, 419–427 (2012).

223. Shimokawa, T. *et al.* Identification of TOMM34, which shows elevated expression in the majority of human colon cancers, as a novel drug target. *Int J Oncol* **29**, 381–386 (2006).

224. Luo, Y. *et al.* Differences in DNA Methylation Signatures Reveal Multiple Pathways of Progression from Adenoma to Colorectal Cancer. *Gastroenterology* **147**, 418–420 (2014).

225. Shirafkan, N. *et al.* microRNA-193a-5p inhibits migration of human HT-29 colon cancer cells via suppression of metastasis pathway. *J Cell Biochem* **120**, 8775–8783 (2018).

226. Dawson, M. A. The cancer epigenome: Concepts, challenges, and therapeutic opportunities. *Science* **355**, 1147–1152 (2017).

227. Wu, Q. *et al.* The BRG1 ATPase of human SWI/SNF chromatin remodeling enzymes as a driver of cancer. *Epigenomics* **9**, 919–931 (2017).

228. Tam, W. L. & Weinberg, R. A. The epigenetics of epithelial-mesenchymal plasticity in cancer. *Nature Medicine* **19**, 1438–1449 (2013).

229. Balch, C., Ramapuram, J. B. & Tiwari, A. K. The epigenomics of embryonic pathway signaling in colorectal cancer. *Frontiers in Pharmacology* **8**, 267–274 (2017).

ANNEXES

Annexes

Annex 1: Significant commonly altered genes by the *ZNF518B* knock-down in DLD1 and HCT116 cell line.

ID	Gene Symbol	Description	Chromosome	Strand
TC0100007789.hg.1	*AGO3*	argonaute RISC catalytic component 3	chr1	+
TC0100009142.hg.1	*PTBP2*	polypyrimidine tract binding protein 2	chr1	+
TC0100010121.hg.1	*HCN3*	hyperpolarization activated cyclic nucleotide gated potassium channel 3	chr1	+
TC0100013483.hg.1	*TAF12*	TAF12 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 20kDa	chr1	-
TC0100014776.hg.1	*DDAH1*	dimethylarginine dimethylaminohydrolase 1	chr1	-
TC0100015872.hg.1	*S100A14*	S100 calcium binding protein A14	chr1	-
TC0200008803.hg.1	*GCC2*	GRIP and coiled-coil domain containing 2	chr2	+
TC0200009065.hg.1	*DBI*	diazepam binding inhibitor (GABA receptor modulator, acyl-CoA binding protein)	chr2	+
TC0200009443.hg.1	*R3HDM1*	Transcript Identified by AceView, Entrez Gene ID(s) 23518	chr2	+
TC0200009687.hg.1	*ARL6IP6*	ADP-ribosylation factor like GTPase 6 interacting protein 6	chr2	+
TC0200009955.hg.1	*CYBRD1*	cytochrome b reductase 1	chr2	+
TC0200010801.hg.1	*RQCD1*	RCD1 required for cell differentiation1 homolog (S. pombe)	chr2	+
TC0200013257.hg.1	*TGOLN2*	trans-golgi network protein 2	chr2	-
TC0200016648.hg.1	*CDC42EP3*	CDC42 effector protein (Rho GTPase binding) 3	chr2	-
TC0300012048.hg.1	*ZBTB20; MIR568*	zinc finger and BTB domain containing 20; microRNA 568	chr3	-
TC0400010048.hg.1	*ZNF518B*	zinc finger protein 518B	chr4	-
TC0400012251.hg.1	*C4orf46*	chromosome 4 open reading frame 46	chr4	-
TC0500009061.hg.1	*PCYOX1L*	prenylcysteine oxidase 1 like	chr5	+
TC0500011984.hg.1	*HINT1*	histidine triad nucleotide binding protein 1	chr5	-
TC0500012210.hg.1	*TMEM173*	transmembrane protein 173	chr5	-
TC0500012486.hg.1	*DCTN4*	dynactin 4 (p62)	chr5	-
TC0600007311.hg.1	*HMGN4*	high mobility group nucleosomal binding domain 4	chr6	+
TC0600011943.hg.1	*ENPP5*	ectonucleotide pyrophosphatase/phosphodiesterase 5 (putative)	chr6	-
TC0600014111.hg.1	*SYNGAP1; MIR5004*	synaptic Ras GTPase activating protein 1; microRNA 5004	chr6	+
TC0600014192.hg.1	*RP11-307P5.1; SAMD5*	novel transcript; Transcript Identified by AceView, Entrez Gene ID(s) 389432	chr6	+
TC0700006618.hg.1	*WIPI2*	WD repeat domain, phosphoinositide interacting 2	chr7	+
TC0700007034.hg.1	*CREB5*	cAMP responsive element binding protein 5	chr7	+
TC0700011318.hg.1	*ERV3-1; ZNF117*	endogenous retrovirus group 3, member 1; zinc finger protein 117	chr7	-
TC0700011710.hg.1	*SLC25A40*	solute carrier family 25, member 40	chr7	-
TC0800010285.hg.1	*SLC20A2*	solute carrier family 20 (phosphate transporter), member 2	chr8	-
TC0900009930.hg.1	*FANCG*	Fanconi anemia complementation group G	chr9	-
TC0900010886.hg.1	*PTCH1*	patched 1	chr9	-
TC0900011432.hg.1	*RBM18*	RNA binding motif protein 18	chr9	-
TC0900011501.hg.1	*NR6A1*	nuclear receptor subfamily 6, group A, member 1	chr9	-
TC0900011655.hg.1	*ZER1*	zyg-11 related, cell cycle regulator	chr9	-
TC0X00007744.hg.1	*SH3BGRL*	SH3 domain binding glutamate-rich protein like	chrX	+
TC0X00008514.hg.1	*DDX26B*	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 26B	chrX	+

TC0X00009007.hg.1	*PUDP*	pseudouridine 5-phosphatase	chrX	-
TC0X00009386.hg.1	*SRPX*	sushi-repeat containing protein, X-linked	chrX	-
TC0X00009650.hg.1	*WDR45; PRAF2*	WD repeat domain 45; PRA1 domain family, member 2	chrX	-
TC1000006796.hg.1	*SEC61A2*	Sec61 translocon alpha 2 subunit	chr10	+
TC1000007272.hg.1	*CREM*	cAMP responsive element modulator	chr10	+
TC1000007954.hg.1	*SLC29A3*	solute carrier family 29 (equilibrative nucleoside transporter), member 3	chr10	+
TC1000010273.hg.1	*NRP1*	neuropilin 1	chr10	-
TC1100009248.hg.1	*RNF26*	ring finger protein 26	chr11	+
TC1200007147.hg.1	*ARNTL2*	aryl hydrocarbon receptor nuclear translocator-like 2	chr12	+
TC1200009734.hg.1	*VAMP1*	vesicle associated membrane protein 1	chr12	-
TC1500006925.hg.1	*THBS1*	thrombospondin 1	chr15	+
TC1500007034.hg.1	*SNAP23*	synaptosome associated protein 23kDa	chr15	+
TC1600007887.hg.1	*RBL2*	retinoblastoma-like 2	chr16	+
TC1700006464.hg.1	*FAM57A*	family with sequence similarity 57, member A	chr17	+
TC1700007102.hg.1	*ALKBH5*	alkB homolog 5, RNA demethylase	chr17	+
TC1700008263.hg.1	*ABCC3*	ATP binding cassette subfamily C member 3	chr17	+
TC1700008882.hg.1	*KCTD2*	potassium channel tetramerization domain containing 2	chr17	+
TC1700008897.hg.1	*MRPS7*	mitochondrial ribosomal protein S7	chr17	+
TC1700011919.hg.1	*CEP295NL; TIMP2*	CEP295 N-terminal like; TIMP metallopeptidase inhibitor 2	chr17	-
TC1900008286.hg.1	*BCAM*	basal cell adhesion molecule (Lutheran blood group)	chr19	+
TC1900009204.hg.1	*MOB3A*	MOB kinase activator 3A	chr19	-
TC1900011194.hg.1	*KCNC3*	potassium channel, voltage gated Shaw related subfamily C, member 3	chr19	-
TC2000009204.hg.1	*TOMM34*	translocase of outer mitochondrial membrane 34	chr20	-
TC2000009218.hg.1	*SDC4*	syndecan 4	chr20	-
TC2000009392.hg.1	*B4GALT5*	UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 5	chr20	-
TC2100008297.hg.1	*SIK1*	salt-inducible kinase 1	chr21	-
TC2200006540.hg.1	*USP18*	ubiquitin specific peptidase 18	chr22	+

Annex 2: Significant altered genes by the *ZNF518B* knock-down in DLD1 cell line. Genes are ordered from higher to lower fold-change.

ID	DLD1 Control Avg (log2)	DLD1 siZNF518B Avg (log2)	Fold Change	P-val	FDR P-val	Gene Symbol	Description
TC0X00009007.hg.1	8.22	6	4.65	3.46E-07	0.0003	*PUDP*	pseudouridine 5-phosphatase
TC2000009204.hg.1	13.17	11.12	4.13	6.55E-09	4.68E-05	*TOMM34*	translocase of outer mitochondrial membrane 34
TC0100007747.hg.1	8.24	6.34	3.72	1.47E-06	0.0005	*GJB3*	gap junction protein beta 3
TC0600008760.hg.1	8.47	6.62	3.6	5.85E-07	0.0003	*PM20D2*	peptidase M20 domain containing 2
TC0900012024.hg.1	6.68	4.85	3.54	1.22E-07	0.0002	*CLIC3*	chloride intracellular channel 3
TC0900007576.hg.1	13.65	11.9	3.37	4.22E-09	4.68E-05	*ANXA1*	annexin A1
TC1200009800.hg.1	12.03	10.39	3.12	2.37E-07	0.0003	*SLC2A3*	solute carrier family 2 (facilitated glucose transporter), member 3
TC0400010048.hg.1	7	5.38	3.08	3.52E-07	0.0003	*ZNF518B*	zinc finger protein 518B
TC1200010788.hg.1	6.89	5.29	3.04	4.97E-07	0.0003	*SPRYD3*	SPRY domain containing 3
TC0600011448.hg.1	7.2	5.6	3.03	3.44E-06	0.0008	*LY6G6C*	lymphocyte antigen 6 complex, locus G6C
TC0200006687.hg.1	8.22	6.63	3.03	8.40E-05	0.0049	*HPCAL1*	hippocalcin-like 1
TC0100012278.hg.1	9.01	7.42	3.01	1.51E-06	0.0005	*SCCPDH*	saccharopine dehydrogenase (putative)
TC2000009218.hg.1	10.99	9.42	2.96	1.93E-08	9.19E-05	*SDC4*	syndecan 4
TC1100008037.hg.1	7.54	5.98	2.93	2.60E-06	0.0007	*OVOL1*	ovo-like zinc finger 1
TC1700007102.hg.1	9.37	7.88	2.81	6.89E-08	0.0001	*ALKBH5*	alkB homolog 5, RNA demethylase
TC0200011093.hg.1	6.45	4.97	2.78	1.04E-06	0.0004	*ALPP*	alkaline phosphatase, placental
TC1200009796.hg.1	12.68	11.22	2.75	1.61E-06	0.0005	*SLC2A14*	solute carrier family 2 (facilitated glucose transporter), member 14
TC0600011134.hg.1	9.88	8.43	2.73	1.27E-05	0.0017	*HIST1H4D*	histone cluster 1, H4d
TC1700006464.hg.1	5.82	4.38	2.72	1.43E-06	0.0005	*FAM57A*	family with sequence similarity 57, member A
TC2200009257.hg.1	7.67	6.22	2.72	3.01E-05	0.0029	*TCN2*	transcobalamin II
TC1000008891.hg.1	11.3	9.87	2.69	2.97E-07	0.0003	*DUSP5*	dual specificity phosphatase 5
TC1200008591.hg.1	9.89	8.46	2.69	7.64E-06	0.0012	*ACTR6*	ARP6 actin-related protein 6 homolog (yeast)
TC0500012388.hg.1	7.62	6.21	2.65	0.0043	0.0548	*DPYSL3*	dihydropyrimidinase-like 3
TC1600007448.hg.1	12.22	10.82	2.63	0.0001	0.0060	*CORO1A*	coronin, actin binding protein, 1A
TC1700011808.hg.1	7.85	6.45	2.63	0.0007	0.0175	*ST6GALNAC2*	ST6 (alpha-N-acetyl-neuraminyl-2,3-beta-galactosyl-1,3)-N-acetylgalactosaminide alpha-2,6-sialyltransferase 2
TC0600011234.hg.1	7.19	5.8	2.62	3.27E-06	0.0008	*HIST1H4L*	histone cluster 1, H4l
TC1600010409.hg.1	6.21	4.82	2.62	4.87E-05	0.0036	*BBS2*	Bardet-Biedl syndrome 2
TC0700008430.hg.1	7.64	6.26	2.59	0.0003	0.0095	*SDHAF3*	succinate dehydrogenase complex assembly factor 3
TC1900011382.hg.1	8.12	6.75	2.58	4.35E-06	0.0009	*MYADM*	Memczak2013 ANTISENSE, CDS, coding, INTERNAL best transcript NM_001020819
TC0100011533.hg.1	6.59	5.23	2.56	5.98E-07	0.0003	*ATF3*	activating transcription factor 3
TC1600011378.hg.1	9.64	8.28	2.56	9.45E-07	0.0004	*MVP; PAGR1*	major vault protein; PAXIP1 associated glutamate-rich protein 1
TC1400009248.hg.1	12	10.64	2.56	3.28E-06	0.0008	*DLGAP5*	discs, large (Drosophila) homolog-associated protein 5
TC0600007613.hg.1	10.52	9.18	2.54	2.52E-07	0.0003	*HSPA1A; HSPA1B*	heat shock 70kDa protein 1A; heat shock 70kDa protein 1B

TC0100015336.hg.1	8.11	6.77	2.54	4.32E-06	0.0009	*BCL2L15*	BCL2-like 15
TC0X00007744.hg.1	8.69	7.35	2.52	6.18E-07	0.0003	*SH3BGRL*	SH3 domain binding glutamate-rich protein like
TC1700008897.hg.1	9.52	8.21	2.49	0.0006	0.0161	*MRPS7*	mitochondrial ribosomal protein S7
TC0200011107.hg.1	10.19	8.88	2.47	0.0011	0.0238	*EFHD1*	EF-hand domain family member D1
TC0600007311.hg.1	7.3	6	2.46	5.13E-07	0.0003	*HMGN4*	high mobility group nucleosomal binding domain 4
TC2000009392.hg.1	10.2	8.91	2.45	4.12E-06	0.0009	*B4GALT5*	UDP-Gal:betaGlcNAc beta 1,4-galactosyltransferase, polypeptide 5
TC1800009222.hg.1	5.11	3.82	2.45	6.54E-05	0.0043	*RNF125*	ring finger protein 125, E3 ubiquitin protein ligase
TC0500012166.hg.1	9.62	8.33	2.44	8.61E-06	0.0013	*CDC25C*	cell division cycle 25C
TC0900009930.hg.1	6.98	5.69	2.44	8.38E-05	0.0049	*FANCG*	Fanconi anemia complementation group G
TC1500007062.hg.1	6.6	5.33	2.42	5.34E-07	0.0003	*CKMT1B; CKMT1A*	creatine kinase, mitochondrial 1B; creatine kinase, mitochondrial 1A
TC1100009248.hg.1	10.67	9.39	2.42	1.05E-06	0.0004	*RNF26*	ring finger protein 26
TC0600011130.hg.1	6.91	5.63	2.42	0.0017	0.0310	*HIST1H4C*	Jeck2013 ANTISENSE, CDS, coding, INTERNAL, OVCODE, OVEXON, UTR3 best transcript NM_003542
TC1500007034.hg.1	11.14	9.88	2.4	4.01E-06	0.0009	*SNAP23*	synaptosome associated protein 23kDa
TC0X00008833.hg.1	8.48	7.21	2.4	5.28E-05	0.0038	*FAM50A*	family with sequence similarity 50, member A
TC1100006796.hg.1	7.07	5.81	2.39	2.03E-05	0.0023	*AKIP1*	A kinase (PRKA) interacting protein 1
TC1700008882.hg.1	8.32	7.07	2.38	2.14E-06	0.0006	*KCTD2*	potassium channel tetramerization domain containing 2
TC1000008798.hg.1	9.6	8.36	2.36	1.60E-05	0.0019	*SFR1*	SWI5-dependent homologous recombination repair protein 1
TC0200015214.hg.1	8.73	7.48	2.36	0.0003	0.0108	*OSGEPL1*	O-sialoglycoprotein endopeptidase-like 1
TC0100013483.hg.1	8.91	7.68	2.35	8.41E-06	0.0013	*TAF12*	TAF12 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 20kDa
TC1600011115.hg.1	7.33	6.11	2.34	0.0003	0.0095	*EMC8*	ER membrane protein complex subunit 8
TC0100014776.hg.1	13.11	11.89	2.33	4.71E-05	0.0036	*DDAH1*	dimethylarginine dimethylaminohydrolase 1
TC0100006982.hg.1	11.12	9.9	2.32	3.38E-05	0.0030	*EFHD2*	EF-hand domain family member D2
TC0500012486.hg.1	12.58	11.37	2.31	4.25E-06	0.0009	*DCTN4*	dynactin 4 (p62)
TC0400012251.hg.1	9.64	8.44	2.3	2.47E-06	0.0007	*C4orf46*	chromosome 4 open reading frame 46
TC0500012842.hg.1	12.91	11.71	2.3	0.0002	0.0071	*DUSP1*	dual specificity phosphatase 1
TC1600010732.hg.1	14.86	13.67	2.29	1.11E-06	0.0004	*NQO1*	NAD(P)H dehydrogenase, quinone 1
TC1900011470.hg.1	8.88	7.69	2.29	1.90E-05	0.0022	*UBE2S*	ubiquitin-conjugating enzyme E2S
TSUnmapped00000400.hg.1	4.76	3.58	2.28	0.0001	0.0057	*SERTAD4*	SERTA domain containing 4
TC0100008101.hg.1	9.86	8.67	2.28	0.0001	0.0067	*KIF2C*	kinesin family member 2C
TC0200015127.hg.1	4.68	3.49	2.28	0.0002	0.0073	*PDE1A*	Transcript Identified by AceView, Entrez Gene ID(s) 5136
TC1700011892.hg.1	13.31	12.13	2.27	1.15E-06	0.0004	*TK1*	thymidine kinase 1, soluble
TC2100008297.hg.1	6.58	5.39	2.27	6.13E-06	0.0011	*SIK1*	salt-inducible kinase 1
TC0500007966.hg.1	9.82	8.64	2.27	2.95E-05	0.0028	*XRCC4*	X-ray repair complementing defective repair in Chinese hamster cells 4
TC0100013293.hg.1	12.55	11.37	2.27	0.0001	0.0060	*ID3*	inhibitor of DNA binding 3, dominant negative helix-loop-helix protein
TC1200008675.hg.1	12.95	11.77	2.26	4.38E-06	0.0009	*TXNRD1*	thioredoxin reductase 1
TC1200007051.hg.1	9.13	7.95	2.25	2.37E-06	0.0006	*PYROXD1*	pyridine nucleotide-disulphide oxidoreductase domain 1
TC1400007706.hg.1	12.44	11.27	2.25	2.35E-05	0.0026	*FOS*	FBJ murine osteosarcoma viral oncogene homolog
TC1000007990.hg.1	9.63	8.46	2.25	0.0002	0.0085	*DDIT4*	DNA damage inducible transcript 4
TC0200016366.hg.1	9.83	8.66	2.25	0.0002	0.0092	*THAP4*	THAP domain containing 4
TC0100009723.hg.1	9.53	8.35	2.25	0.0010	0.0225	*FAM72D*	family with sequence similarity 72, member D

TC0X00009386.hg.1	6.51	5.35	2.24	5.90E-05	0.0041	*SRPX*	sushi-repeat containing protein, X-linked
TC0700006735.hg.1	5.28	4.11	2.24	0.0006	0.0161	*SCIN*	scinderin
TC2100008252.hg.1	7.32	6.16	2.24	0.0007	0.0175	*TMPRSS3*	transmembrane protease, serine 3
TC1100009450.hg.1	10.5	9.35	2.22	0.0006	0.0159	*DCPS*	decapping enzyme, scavenger
TC1500009130.hg.1	5.83	4.67	2.22	0.0128	0.1059	*RHOV*	ras homolog family member V
TC1600007806.hg.1	9.29	8.14	2.21	0.0001	0.0057	*HEATR3*	HEAT repeat containing 3
TC0100018186.hg.1	8.17	7.03	2.2	0.0001	0.0063	*NBL1*	neuroblastoma 1, DAN family BMP antagonist
TC0200009687.hg.1	9.88	8.75	2.2	0.0002	0.0077	*ARL6IP6*	ADP-ribosylation factor like GTPase 6 interacting protein 6
TC0500006730.hg.1	12.77	11.63	2.19	2.12E-06	0.0006	*CCT5*	chaperonin containing TCP1, subunit 5 (epsilon)
TC1600006976.hg.1	7.81	6.68	2.19	4.88E-05	0.0036	*ERCC4*	excision repair cross-complementation group 4
TC0900011432.hg.1	10	8.87	2.19	7.52E-05	0.0047	*RBM18*	RNA binding motif protein 18
TC1700009731.hg.1	5.77	4.64	2.19	0.0001	0.0053	*GAS7*	growth arrest-specific 7
TC0600007263.hg.1	7.64	6.51	2.19	0.0013	0.0274	*HIST1H4A*	histone cluster 1, H4a
TC1100009817.hg.1	10.48	9.36	2.18	8.22E-07	0.0004	*PHLDA2*	pleckstrin homology-like domain, family A, member 2
TC0700011710.hg.1	10.64	9.52	2.18	5.27E-06	0.0010	*SLC25A40*	solute carrier family 25, member 40
TC0600007616.hg.1	11.51	10.39	2.18	1.02E-05	0.0014	*HSPA1B; HSPA1A*	heat shock 70kDa protein 1B; heat shock 70kDa protein 1A
TC0X00009124.hg.1	8.98	7.86	2.18	1.58E-05	0.0019	*PIGA*	phosphatidylinositol glycan anchor biosynthesis class A
TC0200009065.hg.1	12.2	11.08	2.17	2.75E-06	0.0007	*DBI*	diazepam binding inhibitor (GABA receptor modulator, acyl-CoA binding protein)
TC0X00009486.hg.1	5.71	4.59	2.17	5.55E-05	0.0039	*MAOB*	monoamine oxidase B
TC1200011567.hg.1	5	3.89	2.17	0.0013	0.0264	*TMCC3; MIR7844*	transmembrane and coiled-coil domain family 3; microRNA 7844
TC0100017216.hg.1	10.2	9.09	2.16	1.74E-05	0.0020	*NEK2*	NIMA-related kinase 2
TC0800009856.hg.1	7.67	6.56	2.16	0.0008	0.0191	*EGR3*	early growth response 3
TC0300012145.hg.1	5.25	4.14	2.16	0.0033	0.0465	*FSTL1; MIR198*	follistatin like 1; microRNA 198
TC1400010730.hg.1	6.84	5.73	2.15	0.0001	0.0068	*EGLN3*	egl-9 family hypoxia-inducible factor 3
TC1100008652.hg.1	7.43	6.32	2.15	0.0026	0.0402	*TMEM126A*	transmembrane protein 126A
TC1000007272.hg.1	7.29	6.19	2.14	0.0241	0.1540	*CREM*	cAMP responsive element modulator
TC0600013433.hg.1	7.66	6.57	2.13	0.0004	0.0116	*SF3B5*	splicing factor 3b subunit 5
TC0100015332.hg.1	7.72	6.64	2.12	3.07E-05	0.0029	*RSBN1*	round spermatid basic protein 1
TC2200006540.hg.1	7.37	6.29	2.12	4.70E-05	0.0036	*USP18*	ubiquitin specific peptidase 18
TC1100012667.hg.1	5.27	4.18	2.12	8.50E-05	0.0049	*OR8B3*	olfactory receptor, family 8, subfamily B, member 3
TC1600007887.hg.1	10.33	9.26	2.11	8.86E-05	0.0050	*RBL2*	retinoblastoma-like 2
TC2000009884.hg.1	8.76	7.68	2.11	0.0064	0.0693	*AP5S1*	adaptor-related protein complex 5, sigma 1 subunit
TC0600006524.hg.1	5.64	4.57	2.1	2.54E-05	0.0026	*FOXQ1*	forkhead box Q1
TC1200010264.hg.1	9.22	8.15	2.1	3.15E-05	0.0029	*IPO8*	importin 8
TC0500008777.hg.1	10.02	8.95	2.1	3.26E-05	0.0030	*KIF20A*	kinesin family member 20A
TC0500011984.hg.1	11.98	10.92	2.1	3.30E-05	0.0030	*HINT1*	histidine triad nucleotide binding protein 1
TC1900009204.hg.1	8.8	7.73	2.1	3.79E-05	0.0031	*MOB3A*	MOB kinase activator 3A
TC1700010647.hg.1	7.9	6.83	2.1	4.52E-05	0.0035	*KRTAP3-1*	keratin associated protein 3-1
TC1600009063.hg.1	10.37	9.3	2.1	5.07E-05	0.0037	*MRPS34*	mitochondrial ribosomal protein S34
TC0500009490.hg.1	6.69	5.62	2.1	0.0002	0.0090	*BNIP1*	BCL2/adenovirus E1B 19kDa interacting protein 1
TC1200007653.hg.1	12.02	10.95	2.1	0.0002	0.0092	*NR4A1*	nuclear receptor subfamily 4, group A, member 1

TC0200008262.hg.1	7.14	6.08	2.09	6.13E-05	0.0042	*RNF181*	ring finger protein 181
TC1000008482.hg.1	8.76	7.7	2.09	0.0010	0.0232	*CEP55*	centrosomal protein 55kDa
TC0100010152.hg.1	12.69	11.63	2.08	1.29E-05	0.0017	*LMNA*	lamin A/C
TC1100011338.hg.1	9.15	8.1	2.08	2.58E-05	0.0026	*CDK2AP2*	cyclin-dependent kinase 2 associated protein 2
TC0100018233.hg.1	7.5	6.45	2.08	3.01E-05	0.0029	*FPGT*	fucose-1-phosphate guanylyltransferase
TC1300007592.hg.1	10.15	9.09	2.08	3.11E-05	0.0029	*NDFIP2*	Nedd4 family interacting protein 2
TC1200007147.hg.1	9.28	8.23	2.08	3.88E-05	0.0032	*ARNTL2*	aryl hydrocarbon receptor nuclear translocator-like 2
TC0700006618.hg.1	10.08	9.02	2.08	4.09E-05	0.0033	*WIPI2*	WD repeat domain, phosphoinositide interacting 2
TC0X00009018.hg.1	7.45	6.39	2.08	8.14E-05	0.0049	*PNPLA4*	patatin-like phospholipase domain containing 4
TC0800011132.hg.1	6.91	5.85	2.08	8.50E-05	0.0049	*GEM*	GTP binding protein overexpressed in skeletal muscle
TC2000007830.hg.1	8.96	7.9	2.08	0.0001	0.0065	*TFAP2C*	transcription factor AP-2 gamma (activating enhancer binding protein 2 gamma)
TC1900008689.hg.1	11.1	10.05	2.08	0.0002	0.0079	*PPP2R1A*	protein phosphatase 2, regulatory subunit A, alpha
TC0900012189.hg.1	10.86	9.8	2.08	0.0005	0.0148	*TMEM141*	transmembrane protein 141
TC0100017073.hg.1	6.91	5.85	2.08	0.0031	0.0444	*SLC45A3*	solute carrier family 45, member 3
TC0600011138.hg.1	8.35	7.29	2.08	0.0041	0.0530	*HIST1H1D*	histone cluster 1, H1d
TC0900011655.hg.1	8.84	7.79	2.07	6.75E-05	0.0044	*ZER1*	zyg-11 related, cell cycle regulator
TC0500012039.hg.1	7.07	6.02	2.07	0.0008	0.0196	*ZCCHC10*	zinc finger, CCHC domain containing 10
TC0600011349.hg.1	6.45	5.41	2.07	0.0093	0.0853	*RNF39*	ring finger protein 39
TC0X00006799.hg.1	11.52	10.47	2.06	4.75E-06	0.0010	*SAT1*	spermidine/spermine N1-acetyltransferase 1
TC0100017172.hg.1	7.45	6.4	2.06	1.00E-05	0.0014	*IRF6*	interferon regulatory factor 6
TC0600007377.hg.1	13.1	12.06	2.06	2.38E-05	0.0026	*HIST1H2BM*	histone cluster 1, H2bm
TC2000007083.hg.1	12.11	11.06	2.06	4.46E-05	0.0035	*ID1*	inhibitor of DNA binding 1, dominant negative helix-loop-helix protein
TC1200011543.hg.1	12.48	11.44	2.06	9.83E-05	0.0053	*PGAM1*	Homo sapiens phosphoglycerate mutase 1 (brain)
TC0800007362.hg.1	7.22	6.18	2.06	0.0004	0.0129	*PLEKHA2*	pleckstrin homology domain containing, family A (phosphoinositide binding specific) member 2
TC0100015872.hg.1	12.29	11.25	2.06	0.0005	0.0148	*S100A14*	S100 calcium binding protein A14
TC1200006874.hg.1	6.82	5.78	2.06	0.0014	0.0279	*BORCS5*	BLOC-1 related complex subunit 5
TC1500007174.hg.1	10.4	9.36	2.05	1.75E-05	0.0020	*EID1*	EP300 interacting inhibitor of differentiation 1
TC0200013257.hg.1	9.56	8.52	2.05	0.0002	0.0084	*TGOLN2*	trans-golgi network protein 2
TC0X00009650.hg.1	8.09	7.06	2.05	0.0013	0.0264	*WDR45; PRAF2*	WD repeat domain 45; PRA1 domain family, member 2
TC2200008425.hg.1	7.91	6.87	2.05	0.0483	0.2293	*LIF*	leukemia inhibitory factor
TC0300014064.hg.1	10.38	9.35	2.04	3.64E-05	0.0030	*PFN2*	profilin 2
TC1500007067.hg.1	6.93	5.89	2.04	0.0005	0.0136	*CKMT1A*	creatine kinase, mitochondrial 1A
TC0400009525.hg.1	9.18	8.15	2.04	0.0006	0.0169	*ANKRD37*	ankyrin repeat domain 37
TC0700013362.hg.1	9.27	8.25	2.04	0.0078	0.0773	*MRPS17*	mitochondrial ribosomal protein S17
TC1100007003.hg.1	17.29	16.27	2.03	5.64E-06	0.0011	*LDHA*	lactate dehydrogenase A
TC0700012812.hg.1	11.63	10.6	2.03	8.10E-06	0.0013	*MKRN1*	makorin ring finger protein 1
TC0200010801.hg.1	11.62	10.6	2.03	1.64E-05	0.0019	*RQCD1*	RCD1 required for cell differentiation1 homolog (S. pombe)
TC1700012189.hg.1	10.3	9.28	2.03	8.78E-05	0.0050	*SENP3*	SUMO1/sentrin/SMT3 specific peptidase 3
TC0100015943.hg.1	9.84	8.82	2.02	0.0008	0.0207	*DPM3*	dolichyl-phosphate mannosyltransferase polypeptide 3
TC0300013835.hg.1	7.38	6.37	2.02	0.0028	0.0426	*TEX264*	testis expressed 264

TC1700009219.hg.1	7.51	6.49	2.02	0.0035	0.0480	*RAC3*	ras-related C3 botulinum toxin substrate 3 (rho family, small GTP binding protein Rac3)
TC0800007766.hg.1	12.03	11.02	2.01	9.38E-06	0.0014	*RAB2A*	RAB2A, member RAS oncogene family
TC1700009394.hg.1	11.93	10.92	2.01	3.56E-05	0.0030	*PRPF8*	pre-mRNA processing factor 8
TC1200006649.hg.1	13.53	12.53	2.01	9.49E-05	0.0052	*TPI1*	triosephosphate isomerase 1
TC0500007911.hg.1	5.12	4.11	2.01	0.0036	0.0488	*SPZ1*	spermatogenic leucine zipper 1
TC1500009709.hg.1	4.89	7.32	-5.39	3.70E-08	0.0001	*CA12*	carbonic anhydrase XII
TC0200009443.hg.1	6.79	8.86	-4.19	1.89E-07	0.0002	*R3HDM1*	Transcript Identified by AceView, Entrez Gene ID(s) 23518
TC0700013443.hg.1	8.39	10.36	-3.93	5.66E-08	0.0001	*CFTR*	cystic fibrosis transmembrane conductance regulator
TC0100009142.hg.1	8.73	10.69	-3.89	2.26E-08	9.19E-05	*PTBP2*	polypyrimidine tract binding protein 2
TC0700011318.hg.1	5.29	7.25	-3.87	6.92E-07	0.0004	*ERV3-1; ZNF117*	endogenous retrovirus group 3, member 1; zinc finger protein 117
TC0800011595.hg.1	6.04	7.99	-3.86	5.32E-07	0.0003	*EXT1; hunera*	Jeck2013 ALT_ACCEPTOR, ALT_DONOR, coding, INTERNAL, intronic best transcript NM_000127; Transcript Identified by AceView
TC0200007273.hg.1	5.58	7.5	-3.79	2.36E-07	0.0003	*RMDN2*	regulator of microtubule dynamics 2
TC1100007453.hg.1	7.45	9.37	-3.78	5.70E-09	4.68E-05	*MDK*	midkine (neurite growth-promoting factor 2)
TC1000010273.hg.1	9.59	11.48	-3.73	9.24E-07	0.0004	*NRP1*	neuropilin 1
TC0600011615.hg.1	6.04	7.91	-3.67	3.20E-05	0.0029	*DEF6*	Memczak2013 ANTISENSE, CDS, coding, INTERNAL best transcript NM_022047
TC0400006491.hg.1	7.05	8.85	-3.49	3.67E-07	0.0003	*IDUA*	iduronidase, alpha-L-
TC0800007016.hg.1	6.66	8.45	-3.46	0.0001	0.0068	*CHMP7*	Transcript Identified by AceView, Entrez Gene ID(s) 91782
TC1600011501.hg.1	12.88	14.67	-3.45	1.31E-06	0.0005	*NPIPB3*	nuclear pore complex interacting protein family, member B3
TC1900010782.hg.1	7.78	9.53	-3.35	2.57E-08	9.19E-05	*ATP1A3*	ATPase, Na+/K+ transporting, alpha 3 polypeptide
TC0500011334.hg.1	5.48	7.2	-3.29	0.0248	0.1566	*TMEM167A*	Transcript Identified by AceView, Entrez Gene ID(s) 153339
TC0200008516.hg.1	5.63	7.33	-3.27	0.0002	0.0077	*CNNM3*	Zhang2013 ALT_ACCEPTOR, ALT_DONOR, coding, INTERNAL, intronic, OVERLAPTX best transcript NM_017623
TC0100011621.hg.1	10.39	12.08	-3.23	4.58E-08	0.0001	*TGFB2; TGFB2-OT1*	transforming growth factor beta 2; TGFB2 overlapping transcript 1
TC0100016983.hg.1	5.78	7.47	-3.22	6.01E-05	0.0041	*CHI3L1*	chitinase 3-like 1 (cartilage glycoprotein-39)
TC1700008382.hg.1	5.21	6.89	-3.21	0.0002	0.0070	*MSI2*	Memczak2013 ALT_ACCEPTOR, ALT_DONOR, coding, INTERNAL, intronic best transcript NM_170721
TC1600009958.hg.1	12.98	14.66	-3.2	1.70E-07	0.0002	*NPIPB4*	nuclear pore complex interacting protein family, member B4
TC1200010006.hg.1	10.48	12.16	-3.19	2.14E-06	0.0006	*PLBD1*	phospholipase B domain containing 1
TC1700010363.hg.1	5.34	6.97	-3.11	3.20E-07	0.0003	*RP11-466A19.5; MYO1D*	Transcript Identified by AceView, Entrez Gene ID(s) 4642; novel transcript, sense intronic MYO1D
TC1100007366.hg.1	10.62	12.26	-3.11	1.91E-06	0.0006	*TTC17*	tetratricopeptide repeat domain 17
TC0600009240.hg.1	6.14	7.75	-3.06	0.0001	0.0060	*KPNA5*	karyopherin alpha 5 (importin alpha 6)
TC1900008012.hg.1	3.94	5.55	-3.05	8.09E-05	0.0049	*CATSPERG*	catsper channel auxiliary subunit gamma
TC1600011505.hg.1	14.04	15.64	-3.03	4.39E-06	0.0009	*NPIPB4*	nuclear pore complex interacting protein family, member B4
TC0600014111.hg.1	7.92	9.51	-3.02	5.46E-07	0.0003	*SYNGAP1; MIR5004*	synaptic Ras GTPase activating protein 1; microRNA 5004
TC0600007387.hg.1	4.13	5.72	-3.01	8.78E-06	0.0013	*OR2B6*	olfactory receptor, family 2, subfamily B, member 6
TC0200007188.hg.1	6.69	8.27	-2.98	0.0010	0.0223	*BIRC6*	Transcript Identified by AceView, Entrez Gene ID(s) 57448
TC1700008263.hg.1	10.65	12.21	-2.95	7.19E-07	0.0004	*ABCC3*	ATP binding cassette subfamily C member 3
TC0100012460.hg.1	6.33	7.89	-2.95	9.25E-05	0.0051	*UBE2J2*	Transcript Identified by AceView, Entrez Gene ID(s) 118424

TC0400010529.hg.1	6.95	8.51	-2.94	2.29E-05	0.0025	*APBB2*	Jeck2013 ALT_ACCEPTOR, ALT_DONOR, coding, INTERNAL, intronic best transcript NM_004307
TC1600009396.hg.1	3.21	4.77	-2.94	0.0004	0.0127	*TNP2*	transition protein 2 (during histone to protamine replacement)
TC0100018526.hg.1	5.55	7.1	-2.93	1.31E-05	0.0017	*ILDR2*	immunoglobulin-like domain containing receptor 2
TC0900010056.hg.1	6.13	7.67	-2.92	8.04E-06	0.0013	*ANKRD18A; FAM95C*	ankyrin repeat domain 18A; family with sequence similarity 95, member C
TC1900008286.hg.1	8.3	9.85	-2.91	6.28E-07	0.0003	*BCAM*	basal cell adhesion molecule (Lutheran blood group)
TC0600011943.hg.1	9.85	11.38	-2.88	8.02E-08	0.0001	*ENPP5*	ectonucleotide pyrophosphatase/phosphodiesterase 5 (putative)
TC0300013520.hg.1	6.5	8.02	-2.87	5.18E-05	0.0038	*CLDN1*	claudin 1
TC0300006696.hg.1	4.38	5.9	-2.87	0.0003	0.0103	*NR2C2*	Transcript Identified by AceView, Entrez Gene ID(s) 7182
TC0400008972.hg.1	5.2	6.71	-2.86	2.04E-06	0.0006	*DCLK2*	doublecortin-like kinase 2
TC0100010926.hg.1	5.68	7.19	-2.85	0.0016	0.0301	*OCLM*	Transcript Identified by AceView, Entrez Gene ID(s) 10896
TSUnmapped00000178.hg.1	8.2	9.7	-2.84	2.67E-05	0.0027	*SLC16A1*	solute carrier family 16 (monocarboxylate transporter), member 1
TC0600010066.hg.1	8.72	10.21	-2.82	9.71E-08	0.0002	*IGF2R*	insulin-like growth factor 2 receptor
TC1200010185.hg.1	11.22	12.71	-2.81	5.46E-06	0.0010	*ITPR2*	inositol 1,4,5-trisphosphate receptor, type 2
TC0200010448.hg.1	8.55	10.04	-2.81	0.0004	0.0128	*STRADB*	STE20-related kinase adaptor beta
TC1400007774.hg.1	6.37	7.84	-2.77	6.83E-07	0.0004	*NRXN3*	neurexin 3
TC0200016354.hg.1	4.26	5.73	-2.77	0.0024	0.0381	*HDLBP*	Transcript Identified by AceView, Entrez Gene ID(s) 3069
TC1000012144.hg.1	5.78	7.24	-2.76	2.96E-06	0.0007	*CTBP2*	Transcript Identified by AceView, Entrez Gene ID(s) 1488
TC1900011398.hg.1	7.32	8.78	-2.76	1.42E-05	0.0018	*TMC4*	transmembrane channel like 4
TC1200006670.hg.1	6.89	8.35	-2.75	5.17E-06	0.0010	*CLSTN3*	calsyntenin 3
TC1600007030.hg.1	10.97	12.42	-2.73	1.20E-06	0.0004	*NPIPA7; NPIPA8; PKD1P1*	nuclear pore complex interacting protein family, member A7; nuclear pore complex interacting protein family, member A8; polycystic kidney disease 1 (autosomal dominant) pseudogene 1
TC1600008407.hg.1	10.54	11.99	-2.73	3.79E-06	0.0009	*NPIPB15*	nuclear pore complex interacting protein family, member B15
TC0100009149.hg.1	6.83	8.29	-2.73	7.50E-05	0.0047	*PTBP2*	Transcript Identified by AceView, Entrez Gene ID(s) 58155
TC0100015797.hg.1	4.25	5.7	-2.73	0.0052	0.0612	*POGZ*	Transcript Identified by AceView, Entrez Gene ID(s) 23126
TC1900011868.hg.1	7	8.44	-2.72	1.30E-06	0.0005	*EPOR*	erythropoietin receptor
TC0100014769.hg.1	6.88	8.32	-2.71	7.23E-06	0.0012	*MCOLN3*	mucolipin 3
TC1600007353.hg.1	12.48	13.91	-2.7	1.03E-05	0.0014	*NPIPB8*	nuclear pore complex interacting protein family, member B8
TC2200009170.hg.1	5.09	6.52	-2.7	0.0006	0.0158	*ARSA*	arylsulfatase A
TC1900007888.hg.1	6.12	7.55	-2.7	0.0013	0.0267	*APLP1*	amyloid beta (A4) precursor-like protein 1
TC0300013513.hg.1	6.51	7.93	-2.68	9.24E-05	0.0051	*P3H2*	prolyl 3-hydroxylase 2
TC1600009916.hg.1	12.66	14.08	-2.67	7.77E-07	0.0004	*NPIPB11*	nuclear pore complex interacting protein family, member B11
TC1600011364.hg.1	13.09	14.51	-2.67	6.33E-06	0.0011	*NPIPB5*	nuclear pore complex interacting protein family, member B5
TC1100009859.hg.1	7.46	8.88	-2.67	2.00E-05	0.0023	*NUP98*	Transcript Identified by AceView, Entrez Gene ID(s) 4928
TC0200013567.hg.1	8.61	10.02	-2.66	1.58E-07	0.0002	*ANKRD36B*	ankyrin repeat domain 36B
TSUnmapped00000246.hg.1	10.17	11.58	-2.66	9.57E-07	0.0004	*CCDC84*	coiled-coil domain containing 84
TC0100017713.hg.1	10.23	11.64	-2.65	4.79E-08	0.0001	*TTC13*	tetratricopeptide repeat domain 13
TC1300008114.hg.1	7.24	8.64	-2.64	0.0002	0.0081	*CUL4A*	Memczak2013 ALT_ACCEPTOR, ALT_DONOR, coding, INTERNAL, intronic best transcript NM_001008895

TC1700008085.hg.1	5.62	7.02	-2.64	0.0009	0.0213	*STH*	saitohin
TC1900011869.hg.1	9.69	11.07	-2.61	2.55E-07	0.0003	*RGL3*	ral guanine nucleotide dissociation stimulator-like 3
TC0200008534.hg.1	8.34	9.72	-2.61	9.00E-07	0.0004	*ANKRD36*	ankyrin repeat domain 36
TC1800008287.hg.1	7.15	8.54	-2.61	0.0006	0.0169	*ANKRD29*	ankyrin repeat domain 29
TC0100015990.hg.1	6.71	8.09	-2.6	1.16E-06	0.0004	*PAQR6*	progestin and adipoQ receptor family member VI
TC1900011194.hg.1	4.2	5.58	-2.6	1.24E-06	0.0004	*KCNC3*	potassium channel, voltage gated Shaw related subfamily C, member 3
TC0100006486.hg.1	9.09	10.47	-2.6	1.11E-05	0.0015	*AGRN*	agrin
TC0500013322.hg.1	5.68	7.06	-2.59	2.68E-05	0.0027	*NAIP*	NLR family, apoptosis inhibitory protein
TC1900011328.hg.1	14.32	15.69	-2.59	0.0005	0.0150	*ZNF160*	Transcript Identified by AceView, Entrez Gene ID(s) 90338
TC0200015894.hg.1	3.15	4.52	-2.59	0.0061	0.0677	*DOCK10*	Transcript Identified by AceView, Entrez Gene ID(s) 55619
TC1600011465.hg.1	7	8.36	-2.58	2.99E-06	0.0007	*ERVK13-1*	endogenous retrovirus group K13, member 1
TC0400012903.hg.1	5.85	7.21	-2.58	1.49E-05	0.0018	*PROM1*	prominin 1
TC1200010601.hg.1	9.35	10.72	-2.57	1.74E-06	0.0005	*FKBP11; ARF3*	FK506 binding protein 11; ADP-ribosylation factor 3
TC1100010688.hg.1	9.53	10.89	-2.57	4.94E-06	0.0010	*PHF21A*	PHD finger protein 21A
TC1100008985.hg.1	9.38	10.73	-2.55	5.22E-06	0.0010	*ATM*	ATM serine/threonine kinase
TC1600011235.hg.1	7.8	9.16	-2.55	4.74E-05	0.0036	*GALNS*	galactosamine (N-acetyl)-6-sulfatase
TC0200007261.hg.1	11.64	12.99	-2.55	0.0005	0.0154	*QPCT*	glutaminyl-peptide cyclotransferase
TC0100011382.hg.1	4.23	5.58	-2.55	0.0254	0.1586	*DYRK3*	dual specificity tyrosine-(Y)-phosphorylation regulated kinase 3
TC0800008943.hg.1	6.17	7.52	-2.54	6.57E-05	0.0043	*PHF20L1*	Transcript Identified by AceView, Entrez Gene ID(s) 51105
TC1700008261.hg.1	4.15	5.5	-2.54	0.0117	0.1005	*EPN3*	Transcript Identified by AceView, Entrez Gene ID(s) 55040
TC0900010886.hg.1	10.63	11.96	-2.52	9.68E-07	0.0004	*PTCH1*	patched 1
TC0900009679.hg.1	4.22	5.55	-2.52	0.0360	0.1935	*MLLT3*	Jeck2013 ALT_ACCEPTOR, ALT_DONOR, coding, INTERNAL, intronic best transcript NM_004529
TC0200012031.hg.1	3.11	4.42	-2.49	0.0051	0.0609	*HADHA*	Transcript Identified by AceView, Entrez Gene ID(s) 3030
TC0M00006434.hg.1	14.07	15.37	-2.48	3.02E-07	0.0003	*ND2*	MTND2
TC2200007505.hg.1	5.88	7.18	-2.48	4.84E-07	0.0003	*sep-03*	septin 3
TC1000012586.hg.1	7.07	8.38	-2.48	1.62E-06	0.0005	*SEC31B*	SEC31 homolog B, COPII coat complex component
TC0800006873.hg.1	7.48	8.79	-2.48	0.0004	0.0128	*PDGFRL*	platelet-derived growth factor receptor-like
TC0400007542.hg.1	10.92	12.23	-2.48	0.0025	0.0398	*KIT*	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog
TC0300006465.hg.1	6.93	8.24	-2.47	7.35E-06	0.0012	*LRRN1*	leucine rich repeat neuronal 1
TC0600014192.hg.1	6.24	7.55	-2.47	1.75E-05	0.0020	*RP11-307P5.1; SAMD5*	novel transcript; Transcript Identified by AceView, Entrez Gene ID(s) 389432
TC0400008483.hg.1	9.27	10.57	-2.46	2.66E-07	0.0003	*UGT8*	UDP glycosyltransferase 8
TC0400012947.hg.1	3.73	5.03	-2.46	0.0001	0.0057	*GPRIN3*	GPRIN family member 3
TC0200016544.hg.1	5.8	7.09	-2.45	7.28E-05	0.0046	*RIF1*	replication timing regulatory factor 1
TC0800008146.hg.1	5.03	6.32	-2.45	0.0007	0.0183	*WWP1*	Transcript Identified by AceView, Entrez Gene ID(s) 11059
TC0800010285.hg.1	9.4	10.68	-2.43	8.22E-06	0.0013	*SLC20A2*	solute carrier family 20 (phosphate transporter), member 2
TC1600009855.hg.1	12.04	13.32	-2.43	1.21E-05	0.0016	*NPIPB6*	nuclear pore complex interacting protein family, member B6
TC1100009200.hg.1	3.77	5.05	-2.43	3.75E-05	0.0031	*CD3E*	CD3e molecule, epsilon (CD3-TCR complex)
TC1700012296.hg.1	9	10.27	-2.42	7.41E-06	0.0012	*PRKAR1A; ARSG*	protein kinase, cAMP-dependent, regulatory, type I, alpha; arylsulfatase G
TC1200009997.hg.1	5.71	6.99	-2.42	2.19E-05	0.0024	*GRIN2B*	glutamate receptor, ionotropic, N-methyl D-aspartate 2B

TC0100007789.hg.1	9.12	10.39	-2.42	3.33E-05	0.0030	*ago-03*	argonaute RISC catalytic component 3
TC1400010756.hg.1	6.05	7.32	-2.42	0.0004	0.0118	*GPR135*	G protein-coupled receptor 135
TC1600011368.hg.1	5.79	7.06	-2.41	4.26E-06	0.0009	*LAT*	linker for activation of T-cells
TC1400010774.hg.1	6.07	7.34	-2.41	6.46E-05	0.0043	*FOXN3*	forkhead box N3
TC0100016322.hg.1	3.84	5.1	-2.4	4.53E-06	0.0009	*GPR161*	G protein-coupled receptor 161
TC1900010531.hg.1	4.83	6.1	-2.4	6.77E-05	0.0044	*ZNF461*	zinc finger protein 461
TSUnmapped00000313.hg.1	8.02	9.29	-2.4	9.84E-05	0.0053	*CCDC84*	coiled-coil domain containing 84
TC0100010674.hg.1	4.32	5.59	-2.4	0.0360	0.1933	*GPR52*	G protein-coupled receptor 52
TC1200009734.hg.1	7.54	8.8	-2.39	2.02E-06	0.0006	*VAMP1*	vesicle associated membrane protein 1
TC0400008093.hg.1	6.97	8.22	-2.38	1.58E-06	0.0005	*PKD2*	polycystic kidney disease 2 (autosomal dominant)
TC0200012980.hg.1	7.8	9.05	-2.38	1.46E-05	0.0018	*ASPRV1; PCBP1-AS1*	aspartic peptidase, retroviral-like 1; PCBP1 antisense RNA 1
TC1100006444.hg.1	9.04	10.29	-2.38	8.54E-05	0.0049	*ATHL1*	ATH1, acid trehalase-like 1 (yeast)
TC0100018280.hg.1	9.89	11.13	-2.37	3.60E-06	0.0008	*FAM231D; LINC00869*	family with sequence similarity 231, member D; long intergenic non-protein coding RNA 869
TC0700009215.hg.1	3.18	4.43	-2.37	0.0018	0.0327	*EXOC4*	exocyst complex component 4
TC1100006495.hg.1	9.42	10.66	-2.36	3.65E-05	0.0030	*TSPAN4*	tetraspanin 4
TC1200007594.hg.1	7.78	9.02	-2.36	0.0008	0.0194	*ASIC1*	acid sensing ion channel 1
TC2000009360.hg.1	7.59	8.83	-2.36	0.0481	0.2289	*PREX1*	Memczak2013 ALT_ACCEPTOR, ALT_DONOR, coding, INTERNAL, intronic best transcript NM_020820
TC0800007460.hg.1	9.56	10.79	-2.35	5.13E-06	0.0010	*HGSNAT*	heparan-alpha-glucosaminide N-acetyltransferase
TC0600010936.hg.1	4.8	6.03	-2.35	4.64E-05	0.0036	*ATXN1*	Memczak2013 ALT_ACCEPTOR, ALT_DONOR, coding, INTERNAL, intronic best transcript NM_000332
TC1600008977.hg.1	4.27	5.5	-2.35	0.0002	0.0070	*CCDC78*	coiled-coil domain containing 78
TC0700009472.hg.1	7.53	8.76	-2.34	5.89E-05	0.0041	*EPHB6*	EPH receptor B6
TC2100007274.hg.1	3.21	4.44	-2.34	0.0018	0.0325	*NDUFV3*	Transcript Identified by AceView, Entrez Gene ID(s) 4731
TC0X00011338.hg.1	9.22	10.43	-2.33	5.00E-06	0.0010	*ARSD*	arylsulfatase D
TC1700012423.hg.1	8.14	9.37	-2.33	4.73E-05	0.0036	*LRRC37A4P*	leucine rich repeat containing 37, member A4, pseudogene
TC0700009696.hg.1	6.98	8.2	-2.33	0.0014	0.0281	*CHPF2; MIR671*	chondroitin polymerizing factor 2; microRNA 671
TC14_GL000194v1_random00006432.hg.1	9.38	10.6	-2.33	0.0063	0.0691	*MAFIP; TEKT4P2*	MAFF interacting protein (pseudogene); tektin 4 pseudogene 2
TC1600010740.hg.1	10.01	11.22	-2.32	6.73E-06	0.0012	*PDXDC2P*	pyridoxal-dependent decarboxylase domain containing 2, pseudogene
TC0100010121.hg.1	6.72	7.94	-2.32	7.38E-06	0.0012	*HCN3*	hyperpolarization activated cyclic nucleotide gated potassium channel 3
TC1400010390.hg.1	7.75	8.97	-2.32	3.10E-05	0.0029	*AHNAK2*	AHNAK nucleoprotein 2
TC1500007238.hg.1	4.13	5.34	-2.32	0.0004	0.0135	*TMOD2*	tropomodulin 2 (neuronal)
TSUnmapped00000154.hg.1	9.21	10.42	-2.31	1.82E-06	0.0005	*LRP6*	LDL receptor related protein 6
TC1100006485.hg.1	10.33	11.54	-2.31	2.79E-06	0.0007	*TMEM80*	transmembrane protein 80
TC2000008381.hg.1	9.29	10.5	-2.31	7.30E-06	0.0012	*JAG1*	jagged 1
TC0100013205.hg.1	10.61	11.82	-2.31	9.92E-06	0.0014	*ECE1*	endothelin converting enzyme 1
TC0500013282.hg.1	5.39	6.6	-2.31	1.14E-05	0.0015	*ZDHHC11*	zinc finger, DHHC-type containing 11
TC0200009955.hg.1	6.98	8.19	-2.31	4.51E-05	0.0035	*CYBRD1*	cytochrome b reductase 1
TC1500008616.hg.1	8.93	10.14	-2.31	0.0001	0.0057	*HERC2P3*	hect domain and RLD 2 pseudogene 3
TC1600007368.hg.1	6.84	8.04	-2.31	0.0002	0.0078	*ATP2A1*	ATPase, Ca++ transporting, cardiac muscle, fast twitch 1

TC1100010207.hg.1	3.93	5.13	-2.31	0.0004	0.0131	*SOX6; MIR6073*	SRY box 6; microRNA 6073
TC0900011501.hg.1	6.03	7.24	-2.31	0.0008	0.0206	*NR6A1*	nuclear receptor subfamily 6, group A, member 1
TC0600014074.hg.1	7.75	8.95	-2.3	8.02E-07	0.0004	*HIVEP1*	human immunodeficiency virus type I enhancer binding protein 1
TC1800008235.hg.1	4.46	5.66	-2.3	0.0010	0.0233	*ABHD3*	Transcript Identified by AceView, Entrez Gene ID(s) 171586
TC0X00008514.hg.1	5.37	6.57	-2.3	0.0011	0.0236	*DDX26B*	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 26B
TC1600011365.hg.1	10.79	11.98	-2.29	4.81E-06	0.0010	*NPIPB9*	nuclear pore complex interacting protein family, member B9
TC0600014268.hg.1	5.73	6.93	-2.29	8.86E-06	0.0013	*PRRT1*	proline-rich transmembrane protein 1
TC1300009522.hg.1	9.11	10.3	-2.29	8.36E-05	0.0049	*UGGT2*	UDP-glucose glycoprotein glucosyltransferase 2
TC1300008927.hg.1	5.5	6.69	-2.29	0.0006	0.0156	*KPNA3*	Transcript Identified by AceView, Entrez Gene ID(s) 3839
TC0100013897.hg.1	9	10.19	-2.28	5.64E-07	0.0003	*P3H1*	prolyl 3-hydroxylase 1
TC1600008646.hg.1	8.02	9.21	-2.28	8.85E-05	0.0050	*ATP2C2*	ATPase, Ca++ transporting, type 2C, member 2
TC1600006826.hg.1	4.27	5.46	-2.28	0.0061	0.0677	*ABAT*	4-aminobutyrate aminotransferase
TC1100009233.hg.1	7.39	8.58	-2.27	5.76E-06	0.0011	*CCDC84*	coiled-coil domain containing 84
TC0200007999.hg.1	7.58	8.77	-2.27	2.94E-05	0.0028	*ZNF638*	Transcript Identified by AceView, Entrez Gene ID(s) 27332
TC0100008100.hg.1	4.23	5.41	-2.27	0.0006	0.0169	*C1orf228*	chromosome 1 open reading frame 228
TC0300008316.hg.1	4.91	6.09	-2.26	7.32E-06	0.0012	*PVRL3*	poliovirus receptor-related 3
TC1300008705.hg.1	4.24	5.41	-2.25	2.46E-05	0.0026	*ELF1*	Transcript Identified by AceView, Entrez Gene ID(s) 1997; 100128628
TC0700013578.hg.1	10.33	11.49	-2.24	9.69E-05	0.0052	*SEMA3A*	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3A
TC0200007194.hg.1	5.23	6.38	-2.23	0.0001	0.0066	*BIRC6*	Transcript Identified by AceView, Entrez Gene ID(s) 57448
TSUnmapped00000073.hg.1	7.32	8.48	-2.23	0.0292	0.1715	*NDUFA10*	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 10, 42kDa
TC1600007037.hg.1	12.92	14.07	-2.22	8.83E-07	0.0004	*NPIPA7*	nuclear pore complex interacting protein family, member A7
TC2000009317.hg.1	5.14	6.29	-2.22	3.54E-05	0.0030	*SULF2*	sulfatase 2
TC0900012173.hg.1	6.87	8.02	-2.21	5.23E-06	0.0010	*GARNL3*	GTPase activating Rap/RanGAP domain-like 3
TC0100014660.hg.1	9.97	11.12	-2.21	0.0001	0.0066	*FUBP1*	far upstream element (FUSE) binding protein 1
TC1500006925.hg.1	11.61	12.75	-2.21	0.0003	0.0099	*THBS1*	thrombospondin 1
TC1900008188.hg.1	7.16	8.3	-2.21	0.0009	0.0213	*MEGF8*	multiple EGF-like-domains 8
TC1100010643.hg.1	5.45	6.59	-2.21	0.0074	0.0754	*TP53I11*	Memczak2013 ALT_ACCEPTOR, ALT_DONOR, coding, INTERNAL, intronic best transcript NM_001076787
TC1900011919.hg.1	7.64	8.78	-2.2	5.83E-05	0.0041	*ZNF708*	zinc finger protein 708
TC0900008250.hg.1	8	9.14	-2.2	0.0001	0.0068	*ZNF189*	zinc finger protein 189
TC1000012130.hg.1	6.12	7.26	-2.2	0.0146	0.1144	*FAM53B*	Memczak2013 ALT_ACCEPTOR, ALT_DONOR, coding, INTERNAL, intronic best transcript NM_014661
TC0500012210.hg.1	10.22	11.35	-2.19	1.75E-06	0.0005	*TMEM173*	transmembrane protein 173
TC1100013057.hg.1	11.1	12.23	-2.19	4.15E-05	0.0033	*TCIRG1*	T-cell, immune regulator 1, ATPase, H+ transporting, lysosomal V0 subunit A3
TC0300013913.hg.1	5.48	6.61	-2.19	0.0124	0.1040	*TPRG1*	tumor protein p63 regulated 1
TC0200010729.hg.1	7.05	8.18	-2.18	3.79E-05	0.0031	*XRCC5*	Transcript Identified by AceView, Entrez Gene ID(s) 7520
TC0100016135.hg.1	7.4	8.52	-2.18	0.0002	0.0090	*SLAMF6*	SLAM family member 6
TC0600009697.hg.1	3.1	4.22	-2.18	0.0022	0.0368	*PHACTR2*	Jeck2013 ALT_ACCEPTOR, ALT_DONOR, coding, INTERNAL, intronic best transcript NM_001100164
TC0100013630.hg.1	5.28	6.4	-2.18	0.0029	0.0429	*YARS*	Transcript Identified by AceView, Entrez Gene ID(s) 8565

TC0700007034.hg.1	3.85	4.98	-2.18	0.0087	0.0829	*CREB5*	cAMP responsive element binding protein 5
TC1600009524.hg.1	11.09	12.21	-2.17	5.98E-06	0.0011	*NPIPA5*	nuclear pore complex interacting protein family, member A5
TC0200011195.hg.1	4.94	6.06	-2.17	0.0002	0.0078	*AGAP1; noygloy; torkeybo*	Jeck2013 ALT_ACCEPTOR, ALT_DONOR, coding, INTERNAL, intronic best transcript NM_001037131; Transcript Identified by AceView
TC1900011796.hg.1	6.14	7.26	-2.17	0.0002	0.0081	*NDUFA3*	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 3, 9kDa
TC0700013401.hg.1	5	6.12	-2.17	0.0004	0.0123	*AC007566.10; GATAD1*	Transcript Identified by AceView, Entrez Gene ID(s) 57798; novel transcript, antisense to ERVW-1 and PEX1
TC1000007954.hg.1	8.12	9.24	-2.17	0.0006	0.0157	*SLC29A3*	solute carrier family 29 (equilibrative nucleoside transporter), member 3
TC1200007804.hg.1	9.74	10.86	-2.17	0.0012	0.0253	*METTL7B*	methyltransferase like 7B
TC0800010945.hg.1	5.96	7.08	-2.17	0.0021	0.0357	*SLC10A5*	solute carrier family 10, member 5
TC2200009267.hg.1	4.56	5.68	-2.17	0.0176	0.1291	*GTPBP1*	GTP binding protein 1
TC0200008627.hg.1	10.49	11.6	-2.16	4.77E-06	0.0010	*NPAS2*	neuronal PAS domain protein 2
TC0700012755.hg.1	8.65	9.76	-2.16	6.62E-06	0.0011	*KIAA1549*	KIAA1549
TC1200006786.hg.1	13.63	14.74	-2.16	0.0001	0.0063	*TMEM52B*	transmembrane protein 52B
TC1900011190.hg.1	3.33	4.44	-2.16	0.0032	0.0458	*IZUMO2*	IZUMO family member 2
TC0300007051.hg.1	6.77	7.87	-2.15	2.31E-06	0.0006	*ITGA9*	integrin alpha 9
TSUnmapped00000267.hg.1	10.15	11.25	-2.15	5.68E-06	0.0011	*LRP6*	LDL receptor related protein 6
TC1700008095.hg.1	11.71	12.81	-2.15	1.98E-05	0.0022	*LRRC37A*	leucine rich repeat containing 37A
TC0300007647.hg.1	4.32	5.42	-2.15	0.0005	0.0152	*PTPRG*	Jeck2013 ALT_ACCEPTOR, ALT_DONOR, coding, INTERNAL, intronic best transcript NM_002841
TC0100011091.hg.1	6.76	7.86	-2.15	0.0009	0.0208	*C1orf53*	chromosome 1 open reading frame 53
TC0500013231.hg.1	4.63	5.74	-2.15	0.0078	0.0772	*CATSPER3*	cation channel, sperm associated 3
TC0400012933.hg.1	8.22	9.31	-2.14	1.16E-06	0.0004	*NAAA*	N-acylethanolamine acid amidase
TC0200016648.hg.1	9.52	10.62	-2.14	2.68E-05	0.0027	*CDC42EP3*	CDC42 effector protein (Rho GTPase binding) 3
TC1300007197.hg.1	7.38	8.48	-2.14	0.0002	0.0072	*DLEU1*	deleted in lymphocytic leukemia 1 (non-protein coding)
TC0100012695.hg.1	4.71	5.8	-2.14	0.0266	0.1628	*DNAJC11*	Zhang2013 ALT_ACCEPTOR, ALT_DONOR, coding, INTERNAL, intronic best transcript NM_018198
TC1600011487.hg.1	11.2	12.29	-2.13	9.67E-06	0.0014	*NPIPA8*	nuclear pore complex interacting protein family, member A8
TC1900010037.hg.1	7.62	8.71	-2.13	2.45E-05	0.0026	*SUGP2*	SURP and G-patch domain containing 2
TC0700010989.hg.1	10.95	12.04	-2.13	3.46E-05	0.0030	*TNS3*	tensin 3
TC0200007421.hg.1	3.55	4.64	-2.13	9.65E-05	0.0052	*CAMKMT*	Memczak2013 ALT_ACCEPTOR, ALT_DONOR, coding, INTERNAL, intronic best transcript NM_024766
TC0700011571.hg.1	5.83	6.92	-2.13	0.0006	0.0168	*YWHAG*	Memczak2013 ALT_ACCEPTOR, ALT_DONOR, coding, INTERNAL, intronic best transcript NM_012479
TC0300013601.hg.1	5.11	6.2	-2.13	0.0009	0.0218	*ATP13A3*	Jeck2013 ALT_ACCEPTOR, ALT_DONOR, coding, INTERNAL, intronic best transcript NM_024524
TC0200010511.hg.1	6.35	7.45	-2.13	0.0010	0.0223	*NBEAL1*	neurobeachin like 1
TC0300007474.hg.1	5.87	6.96	-2.13	0.0010	0.0225	*DNAH1*	dynein, axonemal, heavy chain 1
TC1100009994.hg.1	3.53	4.61	-2.12	8.45E-05	0.0049	*OR2D2*	olfactory receptor, family 2, subfamily D, member 2
TC0100014459.hg.1	3.46	4.55	-2.12	0.0006	0.0170	*LINC01359*	long intergenic non-protein coding RNA 1359
TC2200009161.hg.1	8.26	9.34	-2.12	0.0007	0.0174	*LMF2*	lipase maturation factor 2

TC1100012722.hg.1	5.96	7.04	-2.12	0.0143	0.1134	*CDON*	cell adhesion associated, oncogene regulated
TC1100007430.hg.1	14.36	15.44	-2.11	2.56E-06	0.0007	*GYLTL1B*	glycosyltransferase-like 1B
TC1600011355.hg.1	10.56	11.63	-2.11	9.52E-06	0.0014	*NPIPA1*	nuclear pore complex interacting protein family, member A1
TC1400007478.hg.1	8.86	9.93	-2.11	2.24E-05	0.0025	*GPHN*	gephyrin
TC1600011353.hg.1	10.77	11.85	-2.11	3.31E-05	0.0030	*NPIPA3*	nuclear pore complex interacting protein family, member A3
TC1300008668.hg.1	5.92	7	-2.11	3.64E-05	0.0030	*LHFP*	lipoma HMGIC fusion partner
TC0300012048.hg.1	5.95	7.03	-2.11	4.42E-05	0.0035	*ZBTB20; MIR568*	zinc finger and BTB domain containing 20; microRNA 568
TC0200009470.hg.1	7.48	8.55	-2.11	0.0004	0.0117	*HNMT*	histamine N-methyltransferase
TC1900009808.hg.1	7.34	8.41	-2.1	6.40E-05	0.0042	*ADGRL1*	adhesion G protein-coupled receptor L1
TC1600011354.hg.1	10.76	11.83	-2.09	2.76E-05	0.0027	*NPIPA2*	nuclear pore complex interacting protein family, member A2
TC0200007200.hg.1	9.05	10.11	-2.09	2.96E-05	0.0028	*LTBP1*	latent transforming growth factor beta binding protein 1
TC0600011823.hg.1	7.69	8.75	-2.09	3.42E-05	0.0030	*TRERF1*	transcriptional regulating factor 1
TC1100007675.hg.1	6.08	7.14	-2.09	5.50E-05	0.0039	*RTN4RL2*	reticulon 4 receptor-like 2
TC0200007908.hg.1	3.53	4.59	-2.09	0.0008	0.0190	*ANTXR1*	anthrax toxin receptor 1
TC0400008183.hg.1	5.87	6.93	-2.09	0.0021	0.0358	*BMPR1B*	bone morphogenetic protein receptor type IB
TC1800008750.hg.1	4.62	5.68	-2.09	0.0086	0.0822	*TCF4*	transcription factor 4
TC1200012758.hg.1	4.75	5.81	-2.09	0.0186	0.1345	*TAS2R14*	taste receptor, type 2, member 14
TC0700011554.hg.1	10.93	11.98	-2.08	1.05E-06	0.0004	*PMS2P3*	PMS1 homolog 2, mismatch repair system component pseudogene 3
TC1200007535.hg.1	8.91	9.97	-2.08	2.68E-05	0.0027	*CACNB3*	calcium channel, voltage-dependent, beta 3 subunit
TC0400012854.hg.1	9.5	10.55	-2.08	0.0004	0.0122	*TMEM144*	transmembrane protein 144
TC0600013126.hg.1	11.87	12.92	-2.07	6.05E-06	0.0011	*PTPRK*	protein tyrosine phosphatase, receptor type, K
TC0200016319.hg.1	7.68	8.73	-2.07	1.43E-05	0.0018	*ANKMY1*	ankyrin repeat and MYND domain containing 1
TC0600014256.hg.1	9.84	10.89	-2.07	3.32E-05	0.0030	*GABBR1*	gamma-aminobutyric acid (GABA) B receptor, 1
TC0300007117.hg.1	7.98	9.02	-2.07	0.0004	0.0136	*ENTPD3*	ectonucleoside triphosphate diphosphohydrolase 3
TC0600009256.hg.1	8.94	10	-2.07	0.0006	0.0156	*DCBLD1*	discoidin, CUB and LCCL domain containing 1
TC1900007543.hg.1	5.14	6.18	-2.07	0.0015	0.0287	*ZNF257*	zinc finger protein 257
TC1400009579.hg.1	4.05	5.11	-2.07	0.0018	0.0325	*ADAM20*	ADAM metallopeptidase domain 20
TC1200010839.hg.1	8.79	9.84	-2.06	1.05E-05	0.0014	*ITGA5*	integrin alpha 5
TC1100009667.hg.1	6.57	7.61	-2.06	9.09E-05	0.0051	*ANO9*	anoctamin 9
TC0700009065.hg.1	6.43	7.48	-2.06	0.0001	0.0059	*CCDC136*	coiled-coil domain containing 136
TC0200008803.hg.1	11.87	12.91	-2.06	0.0007	0.0171	*GCC2*	GRIP and coiled-coil domain containing 2
TC0200008462.hg.1	4.82	5.87	-2.06	0.0010	0.0234	*PROM2*	prominin 2
TC1700011919.hg.1	10.69	11.73	-2.05	2.68E-06	0.0007	*CEP295NL; TIMP2*	CEP295 N-terminal like; TIMP metallopeptidase inhibitor 2
TC1700007319.hg.1	10.12	11.16	-2.05	1.28E-05	0.0017	*WSB1*	WD repeat and SOCS box containing 1
TC0500009061.hg.1	7.68	8.72	-2.05	2.76E-05	0.0027	*PCYOX1L*	prenylcysteine oxidase 1 like
TC1800006897.hg.1	10.23	11.26	-2.05	3.31E-05	0.0030	*LAMA3*	laminin, alpha 3
TC0500009423.hg.1	7.98	9.02	-2.05	8.46E-05	0.0049	*NPM1*	Zhang2013 ALT_ACCEPTOR, ALT_DONOR, coding, INTERNAL, intronic best transcript NM_002520
TC0200015893.hg.1	3.25	4.29	-2.05	0.0003	0.0108	*DOCK10*	dedicator of cytokinesis 10
TSUnmapped00000725.hg.1	7.33	8.37	-2.05	0.0004	0.0123	*CCDC84*	coiled-coil domain containing 84
TC0800011566.hg.1	6.12	7.16	-2.05	0.0061	0.0680	*RAD21*	Transcript Identified by AceView, Entrez Gene ID(s) 5885
TC0500013247.hg.1	3.21	4.25	-2.05	0.0072	0.0742	*PCDHB14*	protocadherin beta 14

TC0100009101.hg.1	5.18	6.21	-2.05	0.0159	0.1209	*ABCD3*	Transcript Identified by AceView, Entrez Gene ID(s) 5825
TC1700010314.hg.1	3.4	4.44	-2.05	0.0323	0.1817	*EVI2A; EVI2B*	ecotropic viral integration site 2A; ecotropic viral integration site 2B
TC1100011742.hg.1	6.01	7.03	-2.04	4.37E-05	0.0034	*KCTD21*	potassium channel tetramerization domain containing 21
TC0900006583.hg.1	9.17	10.2	-2.04	0.0002	0.0074	*UHRF2*	ubiquitin-like with PHD and ring finger domains 2, E3 ubiquitin protein ligase
TC0300011364.hg.1	7.65	8.68	-2.04	0.0006	0.0161	*CADPS*	Ca++-dependent secretion activator
TC0200009967.hg.1	4.94	5.97	-2.04	0.0270	0.1636	*DLX1*	distal-less homeobox 1
TC0X00007573.hg.1	11.01	12.04	-2.03	1.29E-05	0.0017	*OGT*	O-linked N-acetylglucosamine (GlcNAc) transferase
TC0600011173.hg.1	11.72	12.75	-2.03	1.81E-05	0.0021	*GUSBP2*	glucuronidase, beta pseudogene 2
TC0600007158.hg.1	9.54	10.56	-2.03	2.30E-05	0.0025	*SOX4*	SRY box 4
TC0100009646.hg.1	7.93	8.95	-2.03	0.0003	0.0110	*NBPF26*	neuroblastoma breakpoint family, member 26
TC1200006454.hg.1	5.28	6.3	-2.03	0.0005	0.0138	*WNK1*	Transcript Identified by AceView, Entrez Gene ID(s) 378465; 65125
TC2000008516.hg.1	5.81	6.84	-2.03	0.0109	0.0956	*SNX5*	Zhang2013 ALT_ACCEPTOR, ALT_DONOR, coding, INTERNAL, intronic best transcript NM_152227
TC1100008330.hg.1	6.93	7.95	-2.02	4.00E-06	0.0009	*IL18BP*	interleukin 18 binding protein
TC0100015947.hg.1	14.94	15.95	-2.02	2.95E-05	0.0028	*THBS3*	Salzman2013 ALT_ACCEPTOR, ALT_DONOR, coding, INTERNAL, intronic best transcript NM_007112
TC1000006796.hg.1	8.21	9.22	-2.02	3.41E-05	0.0030	*SEC61A2*	Sec61 translocon alpha 2 subunit
TC2000006736.hg.1	12.08	13.1	-2.02	3.73E-05	0.0031	*SPTLC3*	serine palmitoyltransferase, long chain base subunit 3
TC2200007493.hg.1	14.72	15.74	-2.02	6.08E-05	0.0041	*MEI1*	Salzman2013 ALT_ACCEPTOR, ALT_DONOR, coding, INTERNAL, intronic best transcript NM_152513
TSUnmapped00000401.hg.1	2.99	4	-2.02	0.0005	0.0147	*INPP5D*	inositol polyphosphate-5-phosphatase D
TC1700010700.hg.1	5.57	6.58	-2.02	0.0013	0.0269	*DNAJC7*	Zhang2013 ALT_ACCEPTOR, ALT_DONOR, coding, INTERNAL, intronic, OVERLAPTX, OVEXON best transcript NM_003315
TC0800006975.hg.1	7.25	8.26	-2.01	0.0003	0.0110	*BMP1*	bone morphogenetic protein 1
TC1100011184.hg.1	7.43	8.43	-2.01	0.0004	0.0120	*SF1*	Memczak2013 ALT_ACCEPTOR, ALT_DONOR, coding, INTERNAL, intronic best transcript NM_201998
TSUnmapped00000106.hg.1	8.34	9.35	-2.01	0.0010	0.0231	*LRP6*	LDL receptor related protein 6
TC1200007809.hg.1	5.91	6.92	-2.01	0.0017	0.0315	*GDF11*	growth differentiation factor 11
TC1200011251.hg.1	9.2	8.2	2	1.00E-05	0.0014	*TSPAN8*	tetraspanin 8
TC0600011135.hg.1	8.46	7.45	2	0.0002	0.0082	*HIST1H3D; HIST1H2AD*	histone cluster 1, H3d; histone cluster 1, H2ad
TC0100011397.hg.1	5.96	4.96	2	0.0003	0.0109	*C4BPB*	complement component 4 binding protein, beta
TC0100013674.hg.1	6.17	5.16	2	0.0015	0.0289	*DLGAP3*	discs, large (Drosophila) homolog-associated protein 3
TC0200010253.hg.1	5.03	6.04	-2	5.13E-05	0.0037	*ANKAR*	ankyrin and armadillo repeat containing
TC1700012361.hg.1	5.93	6.93	-2	0.0003	0.0107	*CDRT1*	CMT1A duplicated region transcript 1
TC0600007241.hg.1	4.52	5.52	-2	0.0114	0.0984	*LRRC16A*	Transcript Identified by AceView, Entrez Gene ID(s) 55604

Annex 3: Significant altered genes by the ZNF518B knock-down in HCT116 cell line. Genes are ordered from higher to lower fold-change.

ID	HCT116 Control Avg (log2)	HCT116 siZNF518B Avg (log2)	Fold Change	P-val	FDR P-val	Gene Symbol	Description
TC0100009621.hg.1	11.48	8.49	7.91	8.47E-09	6.73E-06	*PHGDH*	phosphoglycerate dehydrogenase
TC0X00007744.hg.1	11.38	8.84	5.81	1.02E-09	1.56E-06	*SH3BGRL*	SH3 domain binding glutamate-rich protein like
TC0500012486.hg.1	13.76	11.31	5.47	1.48E-10	5.02E-07	*DCTN4*	dynactin 4 (p62)
TC0400010048.hg.1	8.87	6.43	5.45	4.47E-09	4.57E-06	*ZNF518B*	zinc finger protein 518B
TC0100014776.hg.1	9.97	7.58	5.26	5.74E-08	2.16E-05	*DDAH1*	dimethylarginine dimethylaminohydrolase 1
TC1200007147.hg.1	9.39	7.12	4.81	3.36E-08	1.47E-05	*ARNTL2*	aryl hydrocarbon receptor nuclear translocator-like 2
TC1100007785.hg.1	12.49	10.27	4.64	6.14E-11	4.39E-07	*TMEM109*	transmembrane protein 109
TC0800011150.hg.1	9.47	7.32	4.45	6.68E-08	2.39E-05	*TP53INP1*	tumor protein p53 inducible nuclear protein 1
TC2000007456.hg.1	8.79	6.66	4.4	2.01E-07	4.84E-05	*TTPAL*	tocopherol (alpha) transfer protein-like
TC0200008071.hg.1	14.54	12.45	4.27	6.22E-09	5.56E-06	*MTHFD2*	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 2, methenyltetrahydrofolate cyclohydrolase
TC0900006708.hg.1	10.44	8.46	3.95	7.97E-09	6.58E-06	*SNAPC3*	small nuclear RNA activating complex polypeptide 3
TC1000011370.hg.1	8.19	6.27	3.77	2.65E-09	3.32E-06	*ACTA2*	actin, alpha 2, smooth muscle, aorta
TC0X00009650.hg.1	8.66	6.76	3.74	9.63E-08	2.99E-05	*WDR45; PRAF2*	WD repeat domain 45; PRA1 domain family, member 2
TC0900011655.hg.1	8.89	7.03	3.64	7.32E-08	2.45E-05	*ZER1*	zyg-11 related, cell cycle regulator
TC0200009687.hg.1	9.88	8.01	3.64	5.39E-07	9.32E-05	*ARL6IP6*	ADP-ribosylation factor like GTPase 6 interacting protein 6
TC0700012684.hg.1	9.57	7.72	3.6	6.22E-08	2.26E-05	*AKR1B1*	aldo-keto reductase family 1, member B1 (aldose reductase)
TC0900008028.hg.1	11.66	9.83	3.58	5.72E-09	5.45E-06	*MFSD14B*	major facilitator superfamily domain containing 14B
TC0800007351.hg.1	7.99	6.19	3.5	2.93E-08	1.42E-05	*TACC1*	transforming, acidic coiled-coil containing protein 1
TC1700008342.hg.1	9.57	7.76	3.5	1.66E-07	4.22E-05	*PCTP*	phosphatidylcholine transfer protein
TC1700008897.hg.1	10.54	8.74	3.49	2.38E-06	0.0003	*MRPS7*	mitochondrial ribosomal protein S7
TC1700012260.hg.1	11.99	10.19	3.48	5.83E-10	1.14E-06	*COASY*	CoA synthase
TC2000009392.hg.1	12.13	10.33	3.48	6.56E-09	5.63E-06	*B4GALT5*	UDP-Gal:betaGlcNAc beta 1,4-galactosyltransferase, polypeptide 5
TC0100016162.hg.1	9.24	7.45	3.48	1.36E-07	3.74E-05	*B4GALT3*	UDP-Gal:betaGlcNAc beta 1,4-galactosyltransferase, polypeptide 3
TC1700008882.hg.1	9.74	7.96	3.44	2.77E-09	3.32E-06	*KCTD2*	potassium channel tetramerization domain containing 2
TC0500012935.hg.1	11.11	9.33	3.44	1.38E-08	8.55E-06	*KIAA1191*	KIAA1191
TC2000009218.hg.1	11.28	9.51	3.43	1.01E-09	1.56E-06	*SDC4*	syndecan 4
TC1500007034.hg.1	11.14	9.37	3.4	5.32E-08	2.07E-05	*SNAP23*	synaptosome associated protein 23kDa
TC0200007399.hg.1	7.92	6.16	3.38	0.0006	0.0123	*PLEKHH2*	pleckstrin homology domain containing, family H (with MyTH4 domain) member 2
TC1000008733.hg.1	12.23	10.49	3.35	2.99E-08	1.42E-05	*MFSD13A*	major facilitator superfamily domain containing 13A
TC0800007459.hg.1	10.39	8.65	3.34	5.85E-09	5.45E-06	*POMK*	protein-O-mannose kinase

TC1500008341.hg.1	9.49	7.77	3.29	4.52E-07	8.22E-05	*MAN2A2*	mannosidase, alpha, class 2A, member 2
TC2000010026.hg.1	13.11	11.4	3.28	3.97E-09	4.28E-06	*TMEM189*	transmembrane protein 189
TC0800006980.hg.1	11.34	9.63	3.27	2.67E-08	1.33E-05	*SLC39A14*	solute carrier family 39 (zinc transporter), member 14
TC0200009065.hg.1	11.71	10.01	3.25	1.03E-08	7.86E-06	*DBI*	diazepam binding inhibitor (GABA receptor modulator, acyl-CoA binding protein)
TC0800008416.hg.1	8.87	7.18	3.23	1.17E-06	0.0002	*GRHL2*	grainyhead-like transcription factor 2
TC0X00010966.hg.1	9.47	7.78	3.22	2.47E-07	5.58E-05	*ATP11C*	ATPase, class VI, type 11C
TC0100015866.hg.1	9.05	7.37	3.21	3.23E-05	0.0017	*S100A4*	S100 calcium binding protein A4
TC1200009252.hg.1	9.55	7.87	3.2	5.36E-07	9.32E-05	*TCTN2*	tectonic family member 2
TC1900006622.hg.1	9.99	8.32	3.18	9.91E-07	0.0001	*NCLN*	nicalin
TC0700011876.hg.1	12.88	11.21	3.18	0.0003	0.0085	*ASNS*	asparagine synthetase (glutamine-hydrolyzing)
TC0700007993.hg.1	12.85	11.19	3.16	3.99E-09	4.28E-06	*CLDN4*	claudin 4
TC1100007819.hg.1	7.98	6.34	3.12	1.03E-05	0.0007	*MYRF*	myelin regulatory factor
TC1600010711.hg.1	12.05	10.42	3.08	1.70E-08	9.59E-06	*CHTF8*	chromosome transmission fidelity factor 8
TC0600011697.hg.1	7.13	5.55	2.98	6.17E-07	0.0001	*CCDC167*	coiled-coil domain containing 167
TC0200016419.hg.1	9.88	8.31	2.97	4.48E-08	1.89E-05	*GPN1*	GPN-loop GTPase 1
TC0100013818.hg.1	11.05	9.49	2.95	5.41E-08	2.07E-05	*PPT1*	palmitoyl-protein thioesterase 1
TC0900010968.hg.1	9.08	7.52	2.95	1.16E-06	0.0002	*TRIM14*	tripartite motif containing 14
TC2200008832.hg.1	8.93	7.38	2.94	5.97E-07	9.93E-05	*DESI1*	desumoylating isopeptidase 1
TC1200010902.hg.1	8.39	6.84	2.93	7.80E-07	0.0001	*ANKRD52*	ankyrin repeat domain 52
TC0300010209.hg.1	11.13	9.59	2.92	8.01E-07	0.0001	*TADA3*	transcriptional adaptor 3
TC1200008792.hg.1	7.82	6.28	2.91	1.27E-06	0.0002	*UNG*	uracil DNA glycosylase
TC1400010169.hg.1	10.52	9	2.87	8.42E-08	2.78E-05	*SETD3*	SET domain containing 3
TC1500010080.hg.1	9.56	8.04	2.87	8.88E-05	0.0034	*IMP3*	IMP3, U3 small nucleolar ribonucleoprotein
TC1100011032.hg.1	8.88	7.36	2.86	3.38E-07	6.77E-05	*CYB561A3*	cytochrome b561 family, member A3
TC0200016686.hg.1	11.91	10.4	2.86	2.33E-06	0.0003	*TEX261*	testis expressed 261
TC0800007738.hg.1	12.13	10.62	2.85	1.18E-08	8.16E-06	*SDCBP*	syndecan binding protein
TC1500008245.hg.1	10.41	8.9	2.85	1.35E-06	0.0002	*ABHD2*	abhydrolase domain containing 2
TC0900008887.hg.1	8.76	7.25	2.84	5.88E-07	9.86E-05	*TBC1D13*	TBC1 domain family, member 13
TC0100015872.hg.1	9.1	7.6	2.84	4.06E-06	0.0004	*S100A14*	S100 calcium binding protein A14
TC0200008351.hg.1	8.81	7.31	2.83	1.48E-07	3.95E-05	*RPIA*	ribose 5-phosphate isomerase A
TC0100015182.hg.1	9.89	8.39	2.82	1.10E-07	3.32E-05	*TAF13*	TAF13 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 18kDa
TC0800011211.hg.1	7.61	6.13	2.79	0.0009	0.0173	*NIPAL2*	NIPA-like domain containing 2
TC1200012723.hg.1	10.14	8.67	2.78	3.54E-07	6.84E-05	*ZNF664*	zinc finger protein 664
TC0900009930.hg.1	6.59	5.12	2.78	1.59E-05	0.0010	*FANCG*	Fanconi anemia complementation group G
TC2000010027.hg.1	14.13	12.66	2.77	1.40E-08	8.55E-06	*TMEM189-UBE2V1*	TMEM189-UBE2V1 readthrough
TC0100010525.hg.1	10.25	8.78	2.77	4.90E-07	8.68E-05	*SFT2D2*	SFT2 domain containing 2
TC0700009488.hg.1	12.5	11.03	2.76	1.53E-08	9.09E-06	*CASP2*	caspase 2
TC0200010801.hg.1	11.57	10.11	2.75	8.66E-08	2.82E-05	*RQCD1*	RCD1 required for cell differentiation1 homolog (S. pombe)
TC0600011808.hg.1	8.08	6.62	2.75	9.38E-05	0.0035	*MED20*	mediator complex subunit 20
TC0200011690.hg.1	14.31	12.86	2.74	3.49E-07	6.84E-05	*YWHAQ*	tyrosine 3-monooxygenase/tryptophan 5-

							monooxygenase activation protein, theta
TC2200009028.hg.1	9.74	8.29	2.73	3.70E-07	7.02E-05	*CERK*	ceramide kinase
TC0600011996.hg.1	7.43	5.98	2.73	0.0007	0.0147	*MUT*	methylmalonyl-CoA mutase
TC0100008620.hg.1	5.64	4.19	2.72	0.0032	0.0406	*DNAJC6*	DnaJ (Hsp40) homolog, subfamily C, member 6
TC1100007801.hg.1	14.61	13.18	2.71	5.85E-07	9.86E-05	*TMEM138*	transmembrane protein 138
TC0200008501.hg.1	9.96	8.52	2.71	2.00E-05	0.0012	*ITPRIPL1*	inositol 1,4,5-trisphosphate receptor interacting protein-like 1
TC0900011432.hg.1	10.43	9	2.7	4.63E-06	0.0004	*RBM18*	RNA binding motif protein 18
TC0500009481.hg.1	15.66	14.25	2.67	6.31E-05	0.0027	*ATP6V0E1*	ATPase, H+ transporting, lysosomal 9kDa, V0 subunit e1
TC1000012491.hg.1	9.28	7.87	2.65	2.97E-07	6.36E-05	*ENTPD7*	ectonucleoside triphosphate diphosphohydrolase 7
TC0700011054.hg.1	6.73	5.33	2.65	8.52E-07	0.0001	*GRB10*	growth factor receptor bound protein 10
TC0100006619.hg.1	9.6	8.19	2.65	1.69E-05	0.0010	*TPRG1L*	tumor protein p63 regulated 1-like
TC0100015079.hg.1	8.14	6.74	2.65	7.24E-05	0.0030	*EXTL2*	exostosin-like glycosyltransferase 2
TC0600009534.hg.1	7.97	6.57	2.64	3.79E-05	0.0019	*MYB*	v-myb avian myeloblastosis viral oncogene homolog
TC0600008066.hg.1	10.26	8.87	2.63	7.29E-07	0.0001	*KLHDC3*	kelch domain containing 3
TC1400007695.hg.1	11	9.6	2.63	1.58E-06	0.0002	*EIF2B2*	eukaryotic translation initiation factor 2B, subunit 2 beta, 39kDa
TC2200006540.hg.1	6.97	5.58	2.62	1.66E-07	4.22E-05	*USP18*	ubiquitin specific peptidase 18
TC0800012284.hg.1	8.21	6.83	2.62	2.00E-06	0.0002	*EXTL3*	exostosin-like glycosyltransferase 3
TC1100009248.hg.1	10.01	8.63	2.61	1.49E-07	3.95E-05	*RNF26*	ring finger protein 26
TC0100013483.hg.1	10.04	8.65	2.61	1.67E-07	4.22E-05	*TAF12*	TAF12 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 20kDa
TC0100012089.hg.1	8.6	7.22	2.6	5.18E-07	9.10E-05	*GPR137B*	G protein-coupled receptor 137B
TC1800009284.hg.1	10	8.63	2.59	2.07E-07	4.93E-05	*C18orf32*	chromosome 18 open reading frame 32
TC1400006715.hg.1	6.61	5.23	2.59	0.0033	0.0413	*PCK2*	phosphoenolpyruvate carboxykinase 2 (mitochondrial)
TC1800006635.hg.1	9.96	8.6	2.58	2.54E-08	1.29E-05	*TWSG1*	twisted gastrulation BMP signaling modulator 1
TC1000012577.hg.1	10.11	8.75	2.57	1.43E-07	3.88E-05	*LIPA*	lipase A, lysosomal acid, cholesterol esterase
TC1000008400.hg.1	8.69	7.33	2.57	1.26E-06	0.0002	*IFIT1*	interferon-induced protein with tetratricopeptide repeats 1
TC2000007457.hg.1	6.82	5.46	2.57	0.0002	0.0066	*PKIG*	protein kinase (cAMP-dependent, catalytic) inhibitor gamma
TC1100010051.hg.1	10.29	8.93	2.56	3.58E-07	6.86E-05	*TMEM9B*	TMEM9 domain family, member B
TC1400007691.hg.1	10.29	8.94	2.55	2.96E-06	0.0003	*DLST*	dihydrolipoamide S-succinyltransferase (E2 component of 2-oxo-glutarate complex)
TC1200006905.hg.1	8.39	7.04	2.55	1.01E-05	0.0007	*FAM234B*	family with sequence similarity 234, member B
TC1700008328.hg.1	8.52	7.17	2.55	6.09E-05	0.0027	*STXBP4*	syntaxin binding protein 4
TC1800007543.hg.1	8.45	7.1	2.54	0.0001	0.0037	*SERPINB5*	serpin peptidase inhibitor, clade B (ovalbumin), member 5
TC0400012251.hg.1	10.22	8.88	2.53	2.77E-07	6.05E-05	*C4orf46*	chromosome 4 open reading frame 46
TC1700011208.hg.1	12.01	10.68	2.52	8.98E-08	2.87E-05	*TRIM25; MIR3614*	tripartite motif containing 25; microRNA 3614
TC1700007102.hg.1	10.03	8.7	2.52	7.37E-07	0.0001	*ALKBH5*	alkB homolog 5, RNA demethylase
TC1200012083.hg.1	9.03	7.7	2.52	1.97E-06	0.0002	*FBXO21*	F-box protein 21
TC0100018472.hg.1	10.56	9.22	2.52	7.15E-06	0.0006	*SEC22B*	SEC22 homolog B, vesicle trafficking protein (gene/pseudogene)
TC0700007811.hg.1	10.13	8.8	2.51	0.0002	0.0065	*INTS4P2*	integrator complex subunit 4 pseudogene 2
TC0700013534.hg.1	9.48	8.16	2.49	8.85E-07	0.0001	*GGCT*	gamma-glutamylcyclotransferase
TC0300013838.hg.1	5.7	4.39	2.48	8.30E-06	0.0006	*ABHD14A*	abhydrolase domain containing 14A

TC0500009131.hg.1	13.41	12.1	2.47	3.20E-07	6.63E-05	*G3BP1*	GTPase activating protein (SH3 domain) binding protein 1
TC0900009957.hg.1	8.12	6.82	2.47	1.33E-05	0.0009	*HINT2*	histidine triad nucleotide binding protein 2
TC1500008231.hg.1	11.85	10.55	2.47	1.88E-05	0.0011	*AEN*	apoptosis enhancing nuclease
TC1600008259.hg.1	9.79	8.49	2.47	7.96E-05	0.0032	*PDPR*	pyruvate dehydrogenase phosphatase regulatory subunit
TC1200007161.hg.1	10.77	9.47	2.46	2.00E-06	0.0002	*MRPS35*	mitochondrial ribosomal protein S35
TC1600011347.hg.1	7.67	6.37	2.46	9.02E-06	0.0007	*ALG1; NAGPA-AS1*	ALG1, chitobiosyldiphosphodolichol beta-mannosyltransferase; NAGPA antisense RNA 1
TC1900007688.hg.1	10.32	9.03	2.45	1.00E-06	0.0001	*CCNE1*	cyclin E1
TC0400006579.hg.1	11.49	10.2	2.45	1.13E-06	0.0002	*ADD1*	adducin 1 (alpha)
TC1000007272.hg.1	7.29	6	2.45	0.0013	0.0219	*CREM*	cAMP responsive element modulator
TC0700011785.hg.1	9.22	7.94	2.44	1.89E-07	4.66E-05	*CDK6*	cyclin-dependent kinase 6
TC0700012780.hg.1	13.32	12.04	2.44	3.29E-07	6.66E-05	*HIPK2*	homeodomain interacting protein kinase 2
TC1900009639.hg.1	11.01	9.72	2.44	3.01E-05	0.0016	*AP1M2*	adaptor-related protein complex 1, mu 2 subunit
TC0200013096.hg.1	13.19	11.92	2.43	4.20E-07	7.77E-05	*MOB1A*	MOB kinase activator 1A
TC2200008036.hg.1	6.98	5.7	2.43	8.13E-05	0.0032	*USP41*	ubiquitin specific peptidase 41
TC0300007410.hg.1	13.68	12.4	2.42	1.93E-07	4.70E-05	*GNAI2*	guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 2
TC0800006566.hg.1	11	9.73	2.42	3.60E-06	0.0004	*AGPAT5*	1-acylglycerol-3-phosphate O-acyltransferase 5
TC1700006464.hg.1	5.92	4.65	2.42	4.75E-05	0.0022	*FAM57A*	family with sequence similarity 57, member A
TC1600009200.hg.1	11.72	10.45	2.41	2.99E-06	0.0003	*TRAP1*	TNF receptor-associated protein 1
TC0500013224.hg.1	7.33	6.06	2.41	0.0014	0.0226	*SLC27A6*	solute carrier family 27 (fatty acid transporter), member 6
TC0900008927.hg.1	8.32	7.05	2.41	0.0029	0.0382	*NTMT1*	N-terminal Xaa-Pro-Lys N-methyltransferase 1
TC2000007942.hg.1	6.42	5.16	2.4	0.0018	0.0269	*FAM217B*	family with sequence similarity 217, member B
TC2000006504.hg.1	9.51	8.26	2.39	4.27E-06	0.0004	*STK35*	serine/threonine kinase 35
TC0900008219.hg.1	7.32	6.06	2.39	0.0145	0.1166	*NR4A3*	nuclear receptor subfamily 4, group A, member 3
TC0800011137.hg.1	10.04	8.79	2.38	2.51E-06	0.0003	*RAD54B; FSBP*	RAD54 homolog B (S. cerevisiae); fibrinogen silencer binding protein
TC1700007262.hg.1	8.95	7.71	2.38	3.22E-06	0.0003	*MAP2K3*	mitogen-activated protein kinase kinase 3
TC0600006864.hg.1	10.45	9.2	2.38	0.0002	0.0066	*RIOK1*	RIO kinase 1
TC0600008073.hg.1	8.87	7.63	2.36	7.60E-06	0.0006	*SRF*	serum response factor
TC0100012593.hg.1	8.36	7.12	2.36	0.0026	0.0354	*WRAP73*	WD repeat containing, antisense to TP73
TC1200008942.hg.1	10.56	9.33	2.35	8.57E-07	0.0001	*PLBD2*	phospholipase B domain containing 2
TC1100008447.hg.1	6.96	5.73	2.35	2.12E-05	0.0012	*SPCS2*	signal peptidase complex subunit 2
TC1600007235.hg.1	14.96	13.73	2.34	3.11E-07	6.61E-05	*PLK1*	polo-like kinase 1
TC1000008193.hg.1	12.19	10.96	2.34	5.48E-07	9.33E-05	*PPIF*	peptidylprolyl isomerase F
TC0400006706.hg.1	10.61	9.38	2.34	1.39E-06	0.0002	*TBC1D14*	TBC1 domain family, member 14
TC1500009109.hg.1	8.53	7.31	2.34	2.27E-06	0.0003	*C15orf57; MRPL42P5*	chromosome 15 open reading frame 57; mitochondrial ribosomal protein L42 pseudogene 5
TC0100018473.hg.1	7.17	5.95	2.34	2.82E-06	0.0003	*SEC22B*	Salzman2013 ALT_ACCEPTOR, CDS, coding, INTERNAL, OVCODE, OVEXON best transcript NM_004892; Salzman2013 ANNOTATED, CDS, coding, OVCODE, OVEXON, UTR3 best transcript NM_004892
TC1900006924.hg.1	9.57	8.34	2.34	1.45E-05	0.0010	*ZNF317*	zinc finger protein 317

TC0700011710.hg.1	10.06	8.83	2.33	7.85E-07	0.0001	*SLC25A40*	solute carrier family 25, member 40
TC2200008828.hg.1	9.95	8.74	2.32	3.59E-06	0.0004	*PHF5A*	PHD finger protein 5A
TC1600009942.hg.1	13.07	11.85	2.32	6.44E-06	0.0005	*CDIPT*	CDP-diacylglycerol--inositol 3-phosphatidyltransferase
TC0100006865.hg.1	10.1	8.89	2.32	1.35E-05	0.0009	*AGTRAP*	angiotensin II receptor-associated protein
TC1000007876.hg.1	11.97	10.75	2.32	2.13E-05	0.0012	*DDX21*	DEAD (Asp-Glu-Ala-Asp) box helicase 21
TC0900011778.hg.1	7.6	6.39	2.32	0.0002	0.0068	*RAPGEF1*	Rap guanine nucleotide exchange factor 1
TC0200014672.hg.1	13.05	11.84	2.31	5.04E-05	0.0023	*NR4A2*	nuclear receptor subfamily 4, group A, member 2
TC0300010329.hg.1	10.09	8.89	2.3	7.42E-07	0.0001	*NUP210*	nucleoporin 210kDa
TC0600012542.hg.1	10.15	8.95	2.3	2.94E-06	0.0003	*UBE2J1*	ubiquitin-conjugating enzyme E2, J1
TC1700007254.hg.1	8.42	7.22	2.3	1.83E-05	0.0011	*DHRS7B*	dehydrogenase/reductase (SDR family) member 7B
TC0Y00006882.hg.1	15.91	14.71	2.3	3.02E-05	0.0016	*SLC25A6*	solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 6
TC0300013417.hg.1	11.13	9.94	2.29	7.65E-06	0.0006	*LIPH*	lipase, member H
TC0500012599.hg.1	8.83	7.63	2.29	0.0007	0.0140	*ADAM19*	ADAM metallopeptidase domain 19
TC0X00010675.hg.1	10.6	9.41	2.28	5.41E-06	0.0005	*CUL4B*	cullin 4B
TC0300007290.hg.1	8.91	7.72	2.28	2.95E-05	0.0016	*KLHL18*	kelch-like family member 18
TC1100012422.hg.1	6.72	5.52	2.28	0.0010	0.0180	*APOA4*	apolipoprotein A-IV
TC0100016971.hg.1	14.74	13.55	2.27	8.27E-07	0.0001	*ADIPOR1*	adiponectin receptor 1
TC0800012323.hg.1	7.33	6.15	2.27	1.62E-05	0.0010	*CA13*	carbonic anhydrase XIII
TC0600007311.hg.1	6.27	5.1	2.26	1.58E-06	0.0002	*HMGN4*	high mobility group nucleosomal binding domain 4
TC2000008815.hg.1	11.17	9.99	2.26	5.53E-06	0.0005	*BCL2L1*	BCL2-like 1
TC1700009277.hg.1	10.35	9.18	2.25	8.32E-06	0.0006	*TBCD*	tubulin folding cofactor D
TC0100010102.hg.1	10.3	9.13	2.25	1.04E-05	0.0007	*FLAD1*	flavin adenine dinucleotide synthetase 1
TC0500006822.hg.1	7.75	6.58	2.25	0.0002	0.0058	*OTULIN*	OTU deubiquitinase with linear linkage specificity
TC1000008431.hg.1	8.88	7.72	2.24	1.76E-06	0.0002	*PCGF5*	polycomb group ring finger 5
TC1600008226.hg.1	12.67	11.51	2.24	2.91E-06	0.0003	*NIP7*	NIP7, nucleolar pre-rRNA processing protein
TC2000008268.hg.1	7.55	6.38	2.24	2.50E-05	0.0014	*SLC23A2*	solute carrier family 23 (ascorbic acid transporter), member 2
TC0300013866.hg.1	8.59	7.43	2.24	0.0002	0.0064	*ALG1L2*	ALG1, chitobiosyldiphosphodolichol beta-mannosyltransferase-like 2
TC0300008331.hg.1	9.83	8.67	2.23	4.52E-06	0.0004	*ABHD10*	abhydrolase domain containing 10
TC0100012846.hg.1	11.24	10.08	2.23	2.91E-05	0.0016	*DFFA*	DNA fragmentation factor, 45kDa, alpha polypeptide
TC0300009202.hg.1	13.66	12.51	2.22	1.48E-06	0.0002	*MBNL1*	muscleblind like splicing regulator 1
TC1700008886.hg.1	9.3	8.16	2.22	2.99E-05	0.0016	*SLC16A5*	solute carrier family 16 (monocarboxylate transporter), member 5
TC1700009954.hg.1	8.8	7.65	2.22	8.71E-05	0.0033	*TOM1L2*	target of myb1 like 2 membrane trafficking protein
TC1200009416.hg.1	13.05	11.91	2.21	3.01E-06	0.0003	*RAN*	RAN, member RAS oncogene family
TC1300008487.hg.1	12.34	11.2	2.21	2.32E-05	0.0013	*SLC7A1*	solute carrier family 7 (cationic amino acid transporter, y+ system), member 1
TC0900006881.hg.1	7.51	6.37	2.21	0.0028	0.0374	*IFT74*	intraflagellar transport 74
TC0600014140.hg.1	6.73	5.58	2.21	0.0068	0.0702	*PRIM2*	Salzman2013 ANNOTATED, CDS, coding, OVCODE, OVEXON, UTR3 best transcript NM_000947
TC1600007887.hg.1	8.68	7.55	2.2	9.83E-06	0.0007	*RBL2*	retinoblastoma-like 2
TC0600011870.hg.1	12.41	11.27	2.2	1.32E-05	0.0009	*DNPH1*	2-deoxynucleoside 5-phosphate N-hydrolase 1

TC0200013114.hg.1	10.42	9.29	2.2	1.41E-05	0.0009	*M1AP*	meiosis 1 associated protein
TC1900011684.hg.1	10.42	9.28	2.2	0.0002	0.0063	*BABAM1*	BRISC and BRCA1 A complex member 1
TC0900009489.hg.1	6.32	5.19	2.2	0.0042	0.0492	*GLDC*	glycine dehydrogenase (decarboxylating)
TC0100015921.hg.1	9.4	8.28	2.19	6.42E-06	0.0005	*ADAR*	adenosine deaminase, RNA-specific
TC0100013226.hg.1	10.59	9.46	2.19	7.00E-06	0.0006	*USP48*	ubiquitin specific peptidase 48
TC0700006618.hg.1	11.05	9.92	2.19	6.58E-05	0.0028	*WIPI2*	WD repeat domain, phosphoinositide interacting 2
TC0600014348.hg.1	7.12	5.99	2.19	0.0001	0.0046	*NHSL1*	NHS-like 1
TC1100011128.hg.1	10.94	9.81	2.18	3.49E-05	0.0018	*ATL3*	atlastin GTPase 3
TC1100009973.hg.1	10.54	9.41	2.18	0.0003	0.0071	*ARFIP2*	ADP-ribosylation factor interacting protein 2
TC1900009204.hg.1	9.46	8.34	2.17	1.25E-06	0.0002	*MOB3A*	MOB kinase activator 3A
TC1500007518.hg.1	8.29	7.18	2.17	0.0002	0.0054	*RAB8B*	RAB8B, member RAS oncogene family
TC0200014899.hg.1	10.38	9.26	2.16	3.24E-05	0.0017	*METTL8*	methyltransferase like 8
TC0300007566.hg.1	5.17	4.06	2.16	0.0053	0.0589	*SPATA12*	spermatogenesis associated 12
TC1700007918.hg.1	6.16	5.05	2.16	0.0070	0.0711	*AOC2*	amine oxidase, copper containing 2 (retina-specific)
TC1600010181.hg.1	10.48	9.37	2.15	1.88E-06	0.0002	*NETO2*	neuropilin (NRP) and tolloid (TLL)-like 2
TC0700012798.hg.1	6.99	5.88	2.15	7.79E-06	0.0006	*KDM7A*	lysine (K)-specific demethylase 7A
TC1100012681.hg.1	11.18	10.07	2.15	8.10E-06	0.0006	*SIAE*	sialic acid acetylesterase
TC0700011550.hg.1	13.92	12.82	2.15	1.22E-05	0.0009	*POM121C*	POM121 transmembrane nucleoporin C
TC0600014148.hg.1	7.45	6.35	2.15	3.41E-05	0.0017	*CYB5R4*	cytochrome b5 reductase 4
TC1100009824.hg.1	10.83	9.73	2.15	0.0100	0.0907	*CARS*	cysteinyl-tRNA synthetase
TC1900006488.hg.1	15.16	14.06	2.14	1.93E-06	0.0002	*PTBP1; MIR4745*	polypyrimidine tract binding protein 1; microRNA 4745
TC0400007593.hg.1	14.03	12.94	2.14	1.30E-05	0.0009	*SRP72*	signal recognition particle 72kDa
TC0700009337.hg.1	11.03	9.94	2.14	1.55E-05	0.0010	*C7orf55-LUC7L2; LUC7L2; C7orf55*	C7orf55-LUC7L2 readthrough; LUC7-like 2 pre-mRNA splicing factor; chromosome 7 open reading frame 55
TC1600011401.hg.1	11.08	9.98	2.14	3.22E-05	0.0017	*COQ9*	coenzyme Q9
TC1600007022.hg.1	10.87	9.77	2.14	5.86E-05	0.0026	*ABCC1*	ATP binding cassette subfamily C member 1
TC0400011414.hg.1	6.55	5.45	2.14	0.0016	0.0243	*TRMT10A*	tRNA methyltransferase 10A
TC0900008891.hg.1	8.98	7.89	2.13	6.89E-06	0.0006	*LRRC8A*	leucine rich repeat containing 8 family, member A
TC0200013746.hg.1	8.48	7.39	2.13	2.04E-05	0.0012	*TGFBRAP1*	transforming growth factor beta receptor associated protein 1
TC0200016767.hg.1	6.13	5.04	2.13	3.16E-05	0.0017	*MREG*	melanoregulin
TC0500011984.hg.1	12.56	11.47	2.13	0.0002	0.0061	*HINT1*	histidine triad nucleotide binding protein 1
TC0400009746.hg.1	10.18	9.09	2.13	0.0005	0.0109	*SLBP*	stem-loop binding protein
TC0200013257.hg.1	8.44	7.36	2.12	0.0001	0.0040	*TGOLN2*	trans-golgi network protein 2
TC0400011721.hg.1	8.57	7.49	2.12	0.0005	0.0109	*MAD2L1*	MAD2 mitotic arrest deficient-like 1 (yeast)
TC2200007904.hg.1	10.98	9.9	2.12	0.0014	0.0229	*ATP6V1E1*	ATPase, H+ transporting, lysosomal 31kDa, V1 subunit E1
TC0200007590.hg.1	7.17	6.09	2.12	0.0030	0.0388	*CHAC2*	ChaC, cation transport regulator homolog 2 (E. coli)
TC0500012059.hg.1	14.31	13.24	2.11	8.54E-07	0.0001	*VDAC1*	voltage-dependent anion channel 1
TC1600011350.hg.1	10.95	9.87	2.11	1.51E-05	0.0010	*PMM2*	phosphomannomutase 2
TC0100012454.hg.1	10.27	9.19	2.11	1.67E-05	0.0010	*SDF4*	stromal cell derived factor 4
TC1500010878.hg.1	10.8	9.72	2.1	2.35E-06	0.0003	*VWA9*	von Willebrand factor A domain containing 9
TC2100008297.hg.1	8.31	7.24	2.1	9.50E-05	0.0036	*SIK1*	salt-inducible kinase 1
TC0100009176.hg.1	8.09	7.02	2.1	9.60E-05	0.0036	*SNX7*	sorting nexin 7

TC1200006561.hg.1	7.58	6.51	2.1	0.0001	0.0042	*RAD51AP1*	RAD51 associated protein 1
TC1500010660.hg.1	6.35	5.27	2.1	0.0004	0.0089	*PCSK6*	proprotein convertase subtilisin/kexin type 6
TC1200008544.hg.1	10.2	9.13	2.1	0.0008	0.0153	*NEDD1*	neural precursor cell expressed, developmentally down-regulated 1
TC1600006628.hg.1	12.41	11.35	2.09	3.74E-06	0.0004	*KCTD5*	potassium channel tetramerization domain containing 5
TC0700011575.hg.1	10.85	9.79	2.09	1.60E-05	0.0010	*POMZP3*	POM121 and ZP3 fusion
TC0200007363.hg.1	9.43	8.36	2.09	4.25E-05	0.0021	*MTA3*	metastasis associated 1 family member 3
TC0400008331.hg.1	7.53	6.46	2.09	0.0304	0.1845	*NPNT*	nephronectin
TC2000007016.hg.1	10.79	9.73	2.08	9.94E-06	0.0007	*GINS1*	GINS complex subunit 1 (Psf1 homolog)
TC0300008518.hg.1	9.1	8.04	2.08	1.31E-05	0.0009	*GTF2E1*	general transcription factor IIE subunit 1
TC1200010971.hg.1	11.94	10.88	2.08	1.61E-05	0.0010	*DCTN2*	dynactin 2 (p50)
TC1000012490.hg.1	9.77	8.72	2.08	2.25E-05	0.0013	*PI4K2A*	phosphatidylinositol 4-kinase type 2 alpha
TC0200013104.hg.1	10.42	9.36	2.08	2.42E-05	0.0014	*MOGS*	mannosyl-oligosaccharide glucosidase
TC1200009457.hg.1	8.67	7.61	2.08	3.49E-05	0.0018	*ULK1*	unc-51 like autophagy activating kinase 1
TC2000009823.hg.1	8.05	6.99	2.08	0.0006	0.0133	*STMN3*	stathmin-like 3
TC1200008920.hg.1	10.59	9.54	2.07	2.58E-05	0.0014	*OAS3*	2-5-oligoadenylate synthetase 3
TC0400008876.hg.1	8.86	7.81	2.07	0.0003	0.0071	*USP38*	ubiquitin specific peptidase 38
TC1200007706.hg.1	9.26	8.21	2.07	0.0009	0.0163	*MFSD5*	major facilitator superfamily domain containing 5
TC1200012161.hg.1	7.45	6.4	2.07	0.0017	0.0259	*TRIAP1*	TP53 regulated inhibitor of apoptosis 1
TC0100008978.hg.1	8.76	7.71	2.07	0.0037	0.0450	*LRRC8B*	leucine rich repeat containing 8 family, member B
TC2000009965.hg.1	12.34	11.29	2.06	6.25E-06	0.0005	*FKBP1A; MIR6869*	FK506 binding protein 1A; microRNA 6869
TC0X00008908.hg.1	15.93	14.89	2.06	4.14E-05	0.0020	*SLC25A6*	solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 6
TC0900011035.hg.1	8.87	7.83	2.06	6.69E-05	0.0028	*TMEM246*	transmembrane protein 246
TC0800012451.hg.1	16.26	15.23	2.05	3.98E-06	0.0004	*EIF3E*	eukaryotic translation initiation factor 3, subunit E
TC1500009236.hg.1	8.35	7.31	2.05	8.52E-06	0.0007	*MFAP1*	microfibrillar associated protein 1
TC2100008098.hg.1	9.3	8.26	2.05	9.58E-06	0.0007	*DSCR3*	Down syndrome critical region 3
TC0100016625.hg.1	12.61	11.57	2.05	0.0001	0.0040	*GLUL*	glutamate-ammonia ligase
TC1100011606.hg.1	9.67	8.63	2.05	0.0003	0.0071	*C2CD3*	C2 calcium-dependent domain containing 3
TC1100012019.hg.1	6.4	5.37	2.05	0.0011	0.0190	*SESN3*	sestrin 3
TC1100013012.hg.1	9.73	8.7	2.04	1.87E-06	0.0002	*STX3*	syntaxin 3
TC0500007668.hg.1	11.33	10.3	2.04	1.60E-05	0.0010	*CDK7*	cyclin-dependent kinase 7
TC2200007114.hg.1	13.63	12.6	2.04	2.28E-05	0.0013	*YWHAH*	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, eta
TC2200007677.hg.1	7.13	6.1	2.04	0.0005	0.0111	*TBC1D22A*	TBC1 domain family, member 22A
TC1300006503.hg.1	6.39	5.36	2.04	0.0006	0.0133	*IFT88*	intraflagellar transport 88
TC1600008301.hg.1	6	4.97	2.04	0.0059	0.0633	*CALB2*	calbindin 2
TC1600011545.hg.1	5.4	4.37	2.04	0.0137	0.1129	*PLLP*	plasmolipin
TC0500009837.hg.1	6.82	5.8	2.03	6.18E-05	0.0027	*SLC12A7*	solute carrier family 12 (potassium/chloride transporter), member 7
TC0X00008108.hg.1	7.77	6.75	2.03	0.0084	0.0802	*TMEM164*	transmembrane protein 164
TC0300010357.hg.1	8.61	7.59	2.02	4.94E-05	0.0023	*CHCHD4*	coiled-coil-helix-coiled-coil-helix domain containing 4

TC0X00009007.hg.1	8.03	7.02	2.02	0.0004	0.0100	*PUDP*	pseudouridine 5-phosphatase
TC1900007270.hg.1	9.65	8.64	2.02	0.0004	0.0101	*KLF2*	Kruppel-like factor 2
TC0X00008844.hg.1	5.88	4.87	2.02	0.0030	0.0387	*IKBKG*	inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase gamma
TC1100010893.hg.1	7.68	6.67	2.02	0.0072	0.0723	*SLC43A1*	solute carrier family 43 (amino acid system L transporter), member 1
TC0100014918.hg.1	8.7	7.69	2.02	0.0073	0.0732	*SETSIP*	SET-like protein
TC2200008611.hg.1	13.06	12.05	2.01	9.82E-06	0.0007	*EIF3D*	eukaryotic translation initiation factor 3, subunit D
TC1600011329.hg.1	6.48	5.47	2.01	1.52E-05	0.0010	*TBC1D24*	TBC1 domain family, member 24
TC1100013054.hg.1	10.12	9.12	2.01	0.0002	0.0061	*RAD9A*	RAD9 checkpoint clamp component A
TC1400006821.hg.1	6.61	5.6	2.01	0.0107	0.0955	*G2E3*	G2/M-phase specific E3 ubiquitin protein ligase
TC1400009184.hg.1	5.76	4.75	2.01	0.0115	0.1001	*TXNDC16*	thioredoxin domain containing 16
TC2000009204.hg.1	12.93	9.54	10.48	1.38E-12	1.48E-08	*TOMM34*	translocase of outer mitochondrial membrane 34
TC0X00009386.hg.1	8.22	4.88	10.07	1.42E-10	5.02E-07	*SRPX*	sushi-repeat containing protein, X-linked
TC1500010723.hg.1	11.46	8.14	10.03	2.56E-07	5.71E-05	*CHAC1*	ChaC glutathione-specific gamma-glutamylcyclotransferase 1
TC0100016218.hg.1	6.92	9.69	-6.86	2.61E-10	7.00E-07	*RGS5*	regulator of G-protein signaling 5
TC0100016476.hg.1	9.26	11.97	-6.54	9.04E-11	4.85E-07	*KIAA0040*	KIAA0040
TC0500007972.hg.1	10.09	12.35	-4.79	1.31E-09	1.88E-06	*VCAN*	versican
TC0200008058.hg.1	3.38	5.61	-4.7	1.18E-08	8.16E-06	*C2orf78*	chromosome 2 open reading frame 78
TC1200010182.hg.1	6.35	8.45	-4.26	1.06E-08	7.86E-06	*BHLHE41*	basic helix-loop-helix family, member e41
TC1700012433.hg.1	6.38	8.46	-4.23	1.64E-10	5.02E-07	*HOXB6*	homeobox B6
TC0900010886.hg.1	6.96	9.03	-4.21	4.59E-10	9.84E-07	*PTCH1*	patched 1
TC0500007834.hg.1	9.68	11.67	-3.98	3.51E-10	8.36E-07	*F2RL1*	coagulation factor II (thrombin) receptor-like 1
TC0800006864.hg.1	10.04	12.02	-3.96	9.06E-10	1.56E-06	*ZDHHC2*	zinc finger, DHHC-type containing 2
TC0500011418.hg.1	9.48	11.45	-3.91	4.25E-08	1.82E-05	*MEF2C*	myocyte enhancer factor 2C
TC0300009855.hg.1	3.95	5.88	-3.8	5.86E-08	2.17E-05	*IL1RAP*	interleukin 1 receptor accessory protein
TC0400008943.hg.1	5.89	7.82	-3.8	2.53E-05	0.0014	*EDNRA*	endothelin receptor type A
TC0100009373.hg.1	7.47	9.39	-3.78	2.79E-05	0.0015	*SLC6A17*	solute carrier family 6 (neutral amino acid transporter), member 17
TC1800009268.hg.1	9.92	11.83	-3.77	2.79E-09	3.32E-06	*DSC2*	desmocollin 2
TC1800007360.hg.1	5.75	7.66	-3.77	1.53E-07	3.99E-05	*RAB27B*	RAB27B, member RAS oncogene family
TC0600009439.hg.1	7.22	9.13	-3.76	1.24E-08	8.32E-06	*TMEM200A*	transmembrane protein 200A
TC0100011378.hg.1	5.13	7.03	-3.73	6.41E-06	0.0005	*RASSF5*	Ras association (RalGDS/AF-6) domain family member 5
TC0800010382.hg.1	3.9	5.8	-3.73	3.52E-05	0.0018	*SNAI2*	snail family zinc finger 2
TC0200015364.hg.1	9.02	10.91	-3.69	1.58E-08	9.15E-06	*SATB2*	SATB homeobox 2
TC1000009927.hg.1	5.54	7.38	-3.57	9.44E-07	0.0001	*ST8SIA6*	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 6
TC0600011953.hg.1	6.9	8.73	-3.57	6.16E-05	0.0027	*PLA2G7*	phospholipase A2, group VII (platelet-activating factor acetylhydrolase, plasma)
TC0300012048.hg.1	12.21	14.04	-3.56	4.71E-07	8.49E-05	*ZBTB20; MIR568*	zinc finger and BTB domain containing 20; microRNA 568
TC1700010997.hg.1	8.7	10.51	-3.5	3.26E-08	1.47E-05	*HOXB9*	homeobox B9
TC1400008714.hg.1	8.21	10.02	-3.5	3.57E-05	0.0018	*SLC22A17*	solute carrier family 22, member 17
TC1700008263.hg.1	8.84	10.63	-3.46	2.47E-08	1.29E-05	*ABCC3*	ATP binding cassette subfamily C member 3
TC0600009327.hg.1	4.9	6.69	-3.45	3.07E-08	1.43E-05	*PKIB*	protein kinase (cAMP-dependent, catalytic) inhibitor beta

TC1000009414.hg.1	7.46	9.23	-3.42	2.50E-08	1.29E-05	*INPP5A*	inositol polyphosphate-5-phosphatase A
TC0600011957.hg.1	11.44	13.22	-3.42	7.56E-06	0.0006	*ADGRF1*	adhesion G protein-coupled receptor F1
TC1100009068.hg.1	5.48	7.25	-3.41	6.94E-08	2.41E-05	*NCAM1*	neural cell adhesion molecule 1
TC1900009186.hg.1	6.29	8.05	-3.38	7.08E-08	2.41E-05	*ATP8B3*	ATPase, aminophospholipid transporter, class I, type 8B, member 3
TC0700007034.hg.1	7.17	8.93	-3.38	7.06E-05	0.0030	*CREB5*	cAMP responsive element binding protein 5
TC1700012354.hg.1	7.45	9.2	-3.37	2.13E-06	0.0002	*VAMP2*	vesicle associated membrane protein 2
TC1700010625.hg.1	8.49	10.24	-3.36	1.30E-08	8.46E-06	*TNS4*	tensin 4
TC0300013123.hg.1	6.43	8.17	-3.35	9.81E-08	3.00E-05	*PLD1*	phospholipase D1, phosphatidylcholine-specific
TC0600007087.hg.1	10.1	11.84	-3.34	1.13E-07	3.33E-05	*CAP2*	CAP, adenylate cyclase-associated protein, 2 (yeast)
TC0100011148.hg.1	8.28	10	-3.3	9.11E-08	2.87E-05	*CAMSAP2*	calmodulin regulated spectrin-associated protein family, member 2
TC0400011994.hg.1	7.37	9.09	-3.29	2.23E-06	0.0003	*INPP4B*	inositol polyphosphate-4-phosphatase type II B
TC1200012651.hg.1	5.05	6.76	-3.28	4.79E-07	8.57E-05	*KIF5A*	kinesin family member 5A
TC0X00010515.hg.1	5.49	7.2	-3.28	1.52E-05	0.0010	*CAPN6*	calpain 6
TC0200016648.hg.1	7.3	9.01	-3.27	1.31E-05	0.0009	*CDC42EP3*	CDC42 effector protein (Rho GTPase binding) 3
TC1600009162.hg.1	7.72	9.42	-3.24	1.86E-07	4.65E-05	*HCFC1R1*	host cell factor C1 regulator 1 (XPO1 dependent)
TC1000006796.hg.1	7.06	8.75	-3.23	7.04E-08	2.41E-05	*SEC61A2*	Sec61 translocon alpha 2 subunit
TC0400011600.hg.1	9.41	11.09	-3.22	0.0020	0.0288	*PITX2*	paired-like homeodomain 2
TC0800009891.hg.1	4.53	6.19	-3.16	1.22E-07	3.46E-05	*STC1*	stanniocalcin 1
TC0500011648.hg.1	5.2	6.86	-3.15	2.00E-08	1.10E-05	*EFNA5*	ephrin-A5
TC0600007650.hg.1	5.57	7.22	-3.15	1.12E-07	3.33E-05	*HLA-DRA*	major histocompatibility complex, class II, DR alpha
TC0100015596.hg.1	7.79	9.44	-3.13	0.0001	0.0037	*LIX1L*	limb and CNS expressed 1 like
TC1700012110.hg.1	5.93	7.57	-3.11	2.59E-07	5.73E-05	*SECTM1*	secreted and transmembrane 1
TC0500013149.hg.1	7.29	8.93	-3.11	0.0006	0.0133	*SRD5A1*	steroid-5-alpha-reductase, alpha polypeptide 1 (3-oxo-5 alpha-steroid delta 4-dehydrogenase alpha 1)
TC0600014111.hg.1	7.94	9.56	-3.09	3.29E-07	6.66E-05	*SYNGAP1; MIR5004*	synaptic Ras GTPase activating protein 1; microRNA 5004
TC1600010978.hg.1	3.9	5.5	-3.04	1.71E-05	0.0011	*PKD1L2*	polycystic kidney disease 1-like 2 (gene/pseudogene)
TC1600011328.hg.1	10.39	11.99	-3.03	6.37E-06	0.0005	*NPW*	neuropeptide W
TC0100011022.hg.1	11.13	12.72	-3.02	4.26E-07	7.81E-05	*RGS2*	regulator of G-protein signaling 2
TC0300012781.hg.1	7.32	8.91	-3.02	5.81E-06	0.0005	*WWTR1*	WW domain containing transcription regulator 1
TC0X00007176.hg.1	3.58	5.17	-3.01	0.0011	0.0187	*SSX1*	synovial sarcoma, X breakpoint 1
TC1700010652.hg.1	7.71	9.28	-2.98	8.72E-05	0.0033	*KRTAP2-2*	keratin associated protein 2-2
TC0600009019.hg.1	5.38	6.95	-2.97	5.11E-06	0.0005	*SOBP*	sine oculis binding protein homolog
TC1700012282.hg.1	6.75	8.3	-2.93	1.21E-07	3.46E-05	*SPATA20*	spermatogenesis associated 20
TC0500010540.hg.1	6.1	7.64	-2.91	3.57E-06	0.0004	*LIFR*	leukemia inhibitory factor receptor alpha
TC0500012519.hg.1	5.15	6.69	-2.91	5.00E-06	0.0004	*SPARC*	secreted protein, acidic, cysteine-rich (osteonectin)
TC0700011318.hg.1	6.06	7.6	-2.9	7.36E-06	0.0006	*ERV3-1; ZNF117*	endogenous retrovirus group 3, member 1; zinc finger protein 117
TC0200009955.hg.1	7.81	9.33	-2.87	6.75E-07	0.0001	*CYBRD1*	cytochrome b reductase 1
TC1900008286.hg.1	7.96	9.48	-2.87	7.55E-07	0.0001	*BCAM*	basal cell adhesion molecule (Lutheran blood group)
TC2200006883.hg.1	8.45	9.97	-2.87	0.0001	0.0038	*SUSD2*	sushi domain containing 2
TC1300008609.hg.1	3.94	5.46	-2.86	1.77E-06	0.0002	*DCLK1*	doublecortin-like kinase 1

TC1200008028.hg.1	5.54	7.06	-2.86	1.62E-05	0.0010	*MSRB3*	methionine sulfoxide reductase B3
TC0500010876.hg.1	7	8.51	-2.85	1.32E-07	3.69E-05	*SMIM15*	small integral membrane protein 15
TC1500006925.hg.1	6.39	7.9	-2.84	2.36E-05	0.0013	*THBS1*	thrombospondin 1
TC0600011943.hg.1	9.57	11.07	-2.82	5.26E-08	2.07E-05	*ENPP5*	ectonucleotide pyrophosphatase/phosphodiesterase 5 (putative)
TC0600014192.hg.1	4.26	5.75	-2.81	0.0003	0.0078	*RP11-307P5.1; SAMD5*	novel transcript; Transcript Identified by AceView, Entrez Gene ID(s) 389432
TC0300013108.hg.1	9.08	10.56	-2.8	4.98E-08	2.05E-05	*TNIK*	TRAF2 and NCK interacting kinase
TC0600008146.hg.1	7.02	8.5	-2.8	2.38E-07	5.43E-05	*RUNX2*	runt-related transcription factor 2
TC1800007298.hg.1	5.34	6.82	-2.79	2.09E-06	0.0002	*LIPG*	lipase, endothelial
TC1400009481.hg.1	9.17	10.64	-2.78	5.41E-08	2.07E-05	*PLEK2*	pleckstrin 2
TC1700006524.hg.1	6.27	7.74	-2.78	2.24E-07	5.17E-05	*SERPINF1*	serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 1
TC0300013949.hg.1	6.76	8.23	-2.77	2.10E-07	4.95E-05	*SATB1*	SATB homeobox 1
TC0200014694.hg.1	4.47	5.94	-2.77	2.15E-05	0.0013	*ACVR1C*	activin A receptor type IC
TC1700010651.hg.1	8.38	9.85	-2.77	0.0003	0.0072	*KRTAP2-1*	keratin associated protein 2-1
TC0300011259.hg.1	5.47	6.93	-2.76	3.68E-06	0.0004	*ARHGEF3*	Rho guanine nucleotide exchange factor 3
TC0X00010840.hg.1	5.99	7.45	-2.76	6.38E-06	0.0005	*HS6ST2*	heparan sulfate 6-O-sulfotransferase 2
TC1700011558.hg.1	7.23	8.69	-2.76	0.0001	0.0038	*SLC16A6*	solute carrier family 16, member 6
TC1900010005.hg.1	8.07	9.53	-2.75	2.01E-05	0.0012	*RAB3A*	RAB3A, member RAS oncogene family
TC0400010168.hg.1	10.14	11.59	-2.74	3.35E-08	1.47E-05	*LCORL*	ligand dependent nuclear receptor corepressor like
TC1200007931.hg.1	4.35	5.81	-2.74	8.15E-07	0.0001	*SLC16A7*	solute carrier family 16 (monocarboxylate transporter), member 7
TC1200012711.hg.1	8.43	9.87	-2.73	0.0001	0.0047	*LINC00173*	long intergenic non-protein coding RNA 173
TC1900011067.hg.1	5.44	6.88	-2.72	3.52E-07	6.84E-05	*TMEM143*	transmembrane protein 143
TC1100013130.hg.1	5.98	7.42	-2.71	0.0003	0.0075	*IGF2; INS-IGF2*	insulin-like growth factor 2; INS-IGF2 readthrough
TC1400009006.hg.1	7.57	9	-2.7	3.50E-07	6.84E-05	*TRAPPC6B*	trafficking protein particle complex 6B
TC1400009134.hg.1	9.46	10.89	-2.7	2.57E-06	0.0003	*SAV1*	salvador family WW domain containing protein 1
TSUnmapped00000049.hg.1	5.24	6.67	-2.69	6.44E-05	0.0028	*PADI3*	peptidyl arginine deiminase, type III
TC1400007320.hg.1	4.25	5.66	-2.66	7.16E-07	0.0001	*DACT1*	dishevelled-binding antagonist of beta-catenin 1
TC1200009734.hg.1	8.29	9.71	-2.66	1.14E-06	0.0002	*VAMP1*	vesicle associated membrane protein 1
TC1100011833.hg.1	8.08	9.49	-2.66	2.90E-06	0.0003	*SYTL2*	synaptotagmin-like 2
TC0100013796.hg.1	12.14	13.55	-2.65	1.50E-06	0.0002	*PABPC4*	poly(A) binding protein, cytoplasmic 4 (inducible form)
TC1000007761.hg.1	10.15	11.55	-2.65	1.77E-06	0.0002	*ARID5B*	AT rich interactive domain 5B (MRF1-like)
TC0300012980.hg.1	3.81	5.21	-2.65	4.42E-06	0.0004	*SPTSSB*	serine palmitoyltransferase, small subunit B
TC2000009955.hg.1	5.42	6.83	-2.65	0.0004	0.0090	*SLC2A4RG*	SLC2A4 regulator
TC0400009038.hg.1	9.44	10.85	-2.64	2.23E-06	0.0003	*KIAA0922*	KIAA0922
TC1000010273.hg.1	5.68	7.08	-2.64	9.82E-06	0.0007	*NRP1*	neuropilin 1
TC1400009353.hg.1	7.84	9.24	-2.64	6.53E-05	0.0028	*SIX4*	SIX homeobox 4
TC0300012764.hg.1	9.47	10.87	-2.63	3.80E-06	0.0004	*HLTF*	helicase-like transcription factor
TC1900009603.hg.1	8.53	9.93	-2.63	0.0004	0.0089	*OLFM2*	olfactomedin 2
TC0100016833.hg.1	7.37	8.75	-2.62	1.53E-06	0.0002	*ZBTB41*	zinc finger and BTB domain containing 41
TC1100011888.hg.1	11.3	12.69	-2.62	3.18E-05	0.0017	*CTSC*	cathepsin C

TC0X00009057.hg.1	9.64	11.02	-2.6	2.84E-06	0.0003	*MID1*	midline 1
TC1000007954.hg.1	7.46	8.84	-2.6	1.19E-05	0.0008	*SLC29A3*	solute carrier family 29 (equilibrative nucleoside transporter), member 3
TC0600014191.hg.1	4.83	6.19	-2.58	5.44E-07	9.33E-05	*SAMD5*	sterile alpha motif domain containing 5
TC0X00009215.hg.1	9.43	10.8	-2.58	2.45E-06	0.0003	*EIF1AX*	eukaryotic translation initiation factor 1A, X-linked
TC1200010806.hg.1	8.12	9.49	-2.58	9.77E-06	0.0007	*MAP3K12*	mitogen-activated protein kinase kinase kinase 12
TC1100011141.hg.1	4.72	6.08	-2.57	4.83E-05	0.0023	*RCOR2*	REST corepressor 2
TC0600008697.hg.1	10.35	11.7	-2.55	4.26E-05	0.0021	*NT5E*	5-nucleotidase, ecto (CD73)
TC0600009350.hg.1	6.25	7.6	-2.55	7.46E-05	0.0031	*RNF217*	ring finger protein 217
TC0600011945.hg.1	6.17	7.51	-2.54	9.04E-07	0.0001	*RCAN2*	regulator of calcineurin 2
TC2200007489.hg.1	6.46	7.81	-2.54	1.99E-06	0.0002	*C22orf46*	chromosome 22 open reading frame 46
TC2000008242.hg.1	6.42	7.77	-2.54	2.13E-06	0.0002	*RNF24*	ring finger protein 24
TC1100008969.hg.1	5.05	6.39	-2.54	5.72E-05	0.0026	*ELMOD1*	ELMO/CED-12 domain containing 1
TC0900007552.hg.1	7.58	8.91	-2.52	1.17E-07	3.38E-05	*GDA*	guanine deaminase
TC1900006865.hg.1	5.65	6.98	-2.52	7.67E-06	0.0006	*TGFBR3L*	transforming growth factor beta receptor III like
TC1900011084.hg.1	8.6	9.93	-2.51	9.47E-07	0.0001	*DBP*	D site of albumin promoter (albumin D-box) binding protein
TC1200008269.hg.1	5.71	7.03	-2.5	6.42E-05	0.0028	*NAV3*	neuron navigator 3
TC1000011661.hg.1	6.56	7.88	-2.49	4.15E-07	7.74E-05	*KCNIP2*	Kv channel interacting protein 2
TC1000007199.hg.1	6.97	8.28	-2.49	5.48E-05	0.0025	*ZEB1*	zinc finger E-box binding homeobox 1
TC1100007507.hg.1	7.75	9.06	-2.48	4.04E-07	7.60E-05	*PTPRJ*	protein tyrosine phosphatase, receptor type, J
TC0400011263.hg.1	3.92	5.23	-2.48	1.97E-05	0.0012	*SPARCL1*	SPARC like 1
TC0X00008514.hg.1	5.01	6.32	-2.48	0.0002	0.0061	*DDX26B*	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 26B
TC1700007944.hg.1	6.3	7.6	-2.47	7.64E-07	0.0001	*TMEM106A*	transmembrane protein 106A
TC1700012434.hg.1	8.98	10.28	-2.46	6.02E-06	0.0005	*HOXB7*	homeobox B7
TC1700010657.hg.1	3.72	5.01	-2.45	2.92E-05	0.0016	*KRTAP4-12*	keratin associated protein 4-12
TC1400007168.hg.1	6.61	7.9	-2.45	0.0004	0.0090	*GPR137C*	G protein-coupled receptor 137C
TC2200009153.hg.1	6.35	7.64	-2.44	0.0005	0.0119	*DENND6B*	DENN/MADD domain containing 6B
TC1100013131.hg.1	9.07	10.36	-2.44	0.0006	0.0133	*IGF2*	insulin-like growth factor 2
TC1900011996.hg.1	5.96	7.24	-2.43	3.63E-06	0.0004	*CGB8*	chorionic gonadotropin, beta polypeptide 8
TC1000009067.hg.1	3.89	5.16	-2.42	9.15E-07	0.0001	*GRK5*	G protein-coupled receptor kinase 5
TC0100011364.hg.1	8.41	9.68	-2.42	4.82E-06	0.0004	*SRGAP2*	SLIT-ROBO Rho GTPase activating protein 2
TC0700010563.hg.1	12.05	13.33	-2.42	8.70E-06	0.0007	*HOXA7*	homeobox A7
TC1100007266.hg.1	9.23	10.51	-2.42	3.40E-05	0.0017	*PDHX*	pyruvate dehydrogenase complex, component X
TC2200006827.hg.1	7.85	9.12	-2.42	5.05E-05	0.0023	*GNAZ*	guanine nucleotide binding protein (G protein), alpha z polypeptide
TC1500007619.hg.1	7.88	9.15	-2.41	1.81E-05	0.0011	*SMAD6*	SMAD family member 6
TC1700010653.hg.1	10.49	11.76	-2.41	8.52E-05	0.0033	*KRTAP2-3*	keratin associated protein 2-3
TC1200006738.hg.1	4.46	5.72	-2.4	0.0007	0.0135	*KLRG1*	killer cell lectin-like receptor subfamily G, member 1
TC0300010632.hg.1	10.14	11.4	-2.39	2.16E-07	5.05E-05	*OSBPL10*	oxysterol binding protein-like 10
TC1300008983.hg.1	7.88	9.14	-2.39	2.93E-07	6.35E-05	*INTS6*	integrator complex subunit 6
TC0700013371.hg.1	8.1	9.35	-2.39	8.18E-05	0.0032	*ZNF138*	zinc finger protein 138
TC0800010524.hg.1	9.05	10.3	-2.37	3.18E-07	6.63E-05	*IMPAD1*	inositol monophosphatase domain containing 1
TC1700010604.hg.1	11.66	12.9	-2.37	6.07E-06	0.0005	*NR1D1*	nuclear receptor subfamily 1, group D, member 1

TC0100014752.hg.1	12.34	13.59	-2.37	7.23E-06	0.0006	*GNG5*	guanine nucleotide binding protein (G protein), gamma 5
TC0100008594.hg.1	7.93	9.17	-2.37	7.52E-05	0.0031	*ROR1*	receptor tyrosine kinase-like orphan receptor 1
TC0800010282.hg.1	4.28	5.53	-2.37	0.0001	0.0041	*DKK4*	dickkopf WNT signaling pathway inhibitor 4
TC0200007237.hg.1	8.41	9.65	-2.36	5.25E-06	0.0005	*CRIM1*	cysteine rich transmembrane BMP regulator 1 (chordin-like)
TC1900011194.hg.1	4.14	5.38	-2.36	0.0001	0.0038	*KCNC3*	potassium channel, voltage gated Shaw related subfamily C, member 3
TC0X00011279.hg.1	7.57	8.81	-2.36	0.0019	0.0282	*TSPAN7*	tetraspanin 7
TC1000009792.hg.1	9.42	10.66	-2.35	3.22E-07	6.63E-05	*USP6NL*	USP6 N-terminal like
TC0400008450.hg.1	5.21	6.44	-2.35	3.15E-05	0.0017	*ANK2*	ankyrin 2, neuronal
TC1700012276.hg.1	8.62	9.85	-2.35	0.0002	0.0069	*PNPO*	pyridoxamine 5-phosphate oxidase
TC2000009058.hg.1	6.3	7.53	-2.33	8.99E-05	0.0034	*TGM2*	transglutaminase 2
TC1900011153.hg.1	8.09	9.31	-2.33	0.0004	0.0100	*RRAS*	related RAS viral (r-ras) oncogene homolog
TC1100009431.hg.1	4.48	5.69	-2.31	0.0053	0.0585	*PATE4*	prostate and testis expressed 4
TC1000008585.hg.1	6.21	7.41	-2.3	1.64E-05	0.0010	*MARVELD1*	MARVEL domain containing 1
TC1000007641.hg.1	12.41	13.61	-2.3	2.61E-05	0.0014	*DKK1*	dickkopf WNT signaling pathway inhibitor 1
TC0700013531.hg.1	6.46	7.66	-2.3	9.77E-05	0.0036	*HOXA6*	homeobox A6
TC1900010743.hg.1	9.19	10.39	-2.3	0.0006	0.0121	*TGFB1*	transforming growth factor beta 1
TC0200008803.hg.1	7.37	8.57	-2.3	0.0008	0.0149	*GCC2*	GRIP and coiled-coil domain containing 2
TC1400009194.hg.1	8.13	9.33	-2.3	0.0008	0.0158	*FERMT2*	fermitin family member 2
TC1900011231.hg.1	8.86	10.06	-2.29	1.83E-06	0.0002	*KLK7*	kallikrein related peptidase 7
TC1000006802.hg.1	5.45	6.65	-2.29	8.34E-06	0.0006	*CAMK1D*	calcium/calmodulin-dependent protein kinase ID
TC0400011643.hg.1	11.27	12.47	-2.29	9.43E-06	0.0007	*CAMK2D*	calcium/calmodulin-dependent protein kinase II delta
TC1200010252.hg.1	5.13	6.32	-2.29	5.91E-05	0.0026	*TMTC1*	transmembrane and tetratricopeptide repeat containing 1
TC2200009281.hg.1	4.96	6.16	-2.29	0.0088	0.0833	*PNPLA3*	patatin-like phospholipase domain containing 3
TC1200010569.hg.1	10.07	11.26	-2.28	0.0237	0.1585	*ASB8*	ankyrin repeat and SOCS box containing 8
TC0300007572.hg.1	6.81	7.99	-2.27	0.0003	0.0078	*APPL1*	adaptor protein, phosphotyrosine interaction, PH domain and leucine zipper containing 1
TC1700012279.hg.1	4.32	5.5	-2.27	0.0042	0.0499	*HOXB-AS3*	HOXB cluster antisense RNA 3
TC0100016678.hg.1	4.17	5.35	-2.27	0.0051	0.0574	*FAM129A*	family with sequence similarity 129, member A
TC0400007938.hg.1	10.97	12.14	-2.26	3.46E-06	0.0004	*BMP2K*	BMP2 inducible kinase
TC0100017860.hg.1	10.24	11.42	-2.26	3.78E-06	0.0004	*HEATR1*	HEAT repeat containing 1
TC1100013144.hg.1	7.76	8.93	-2.25	1.67E-05	0.0010	*ZBED5*	zinc finger, BED-type containing 5
TC1000011081.hg.1	7.93	9.1	-2.25	6.77E-05	0.0029	*ZNF503*	zinc finger protein 503
TC0900008150.hg.1	6.25	7.42	-2.25	0.0002	0.0056	*TMOD1*	tropomodulin 1
TC1100011602.hg.1	7.88	9.06	-2.25	0.0003	0.0085	*UCP2*	uncoupling protein 2 (mitochondrial, proton carrier)
TC0100012921.hg.1	9.32	10.49	-2.25	0.0040	0.0479	*DHRS3; MIR6730*	dehydrogenase/reductase (SDR family) member 3; microRNA 6730
TC0100016018.hg.1	12.66	13.82	-2.24	2.11E-06	0.0002	*CRABP2*	cellular retinoic acid binding protein 2
TC0100010755.hg.1	8.21	9.38	-2.24	3.16E-05	0.0017	*RALGPS2*	Ral GEF with PH domain and SH3 binding motif 2
TC1200008597.hg.1	10.82	11.99	-2.24	6.55E-05	0.0028	*GAS2L3*	growth arrest-specific 2 like 3
TC1200011786.hg.1	5.69	6.86	-2.24	0.0007	0.0142	*NUAK1*	NUAK family, SNF1-like kinase, 1
TC1000011659.hg.1	8.94	10.1	-2.23	7.47E-07	0.0001	*MGEA5*	meningioma expressed antigen 5 (hyaluronidase)

TC0100010121.hg.1	4.91	6.07	-2.23	1.65E-05	0.0010	*HCN3*	hyperpolarization activated cyclic nucleotide gated potassium channel 3
TC1300008497.hg.1	8.04	9.2	-2.23	0.0005	0.0111	*UBL3*	ubiquitin-like 3
TC0X00010207.hg.1	4.14	5.3	-2.23	0.0083	0.0798	*POF1B*	premature ovarian failure, 1B
TC0400009033.hg.1	8.98	10.14	-2.22	3.05E-06	0.0003	*TRIM2*	tripartite motif containing 2
TC0500009488.hg.1	5.78	6.92	-2.22	3.54E-06	0.0004	*CREBRF*	CREB3 regulatory factor
TC1900010807.hg.1	7.31	8.46	-2.22	1.86E-05	0.0011	*CEACAM1*	carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein)
TC0100015572.hg.1	8.88	10.03	-2.22	1.91E-05	0.0011	*SRGAP2B*	SLIT-ROBO Rho GTPase activating protein 2B
TC0500012210.hg.1	10.22	11.37	-2.21	7.22E-07	0.0001	*TMEM173*	transmembrane protein 173
TC1400008705.hg.1	8.11	9.26	-2.21	2.61E-05	0.0014	*SLC7A8*	solute carrier family 7 (amino acid transporter light chain, L system), member 8
TC0400008879.hg.1	8.33	9.48	-2.21	8.25E-05	0.0032	*GAB1*	GRB2-associated binding protein 1
TC1900010433.hg.1	3.49	4.63	-2.21	0.0310	0.1868	*SCGB2B2*	secretoglobin, family 2B, member 2
TC1600011317.hg.1	13.13	14.27	-2.2	9.46E-06	0.0007	*NME4*	NME/NM23 nucleoside diphosphate kinase 4
TC0800006496.hg.1	4.21	5.35	-2.2	3.36E-05	0.0017	*MYOM2*	myomesin 2
TC1900009431.hg.1	9.64	10.78	-2.2	5.24E-05	0.0024	*SLC25A23*	solute carrier family 25 (mitochondrial carrier; phosphate carrier), member 23
TC1900011029.hg.1	9.39	10.53	-2.2	6.04E-05	0.0027	*MEIS3*	Meis homeobox 3
TC0400011180.hg.1	11.05	12.19	-2.2	0.0002	0.0055	*SCD5*	stearoyl-CoA desaturase 5
TC1200008535.hg.1	8.57	9.71	-2.2	0.0002	0.0061	*ELK3*	ELK3, ETS-domain protein (SRF accessory protein 2)
TC1700012191.hg.1	6.05	7.19	-2.2	0.0003	0.0072	*CD68*	CD68 molecule
TC0100018454.hg.1	12.61	13.74	-2.19	3.21E-06	0.0003	*ARHGAP29*	Rho GTPase activating protein 29
TC2100008562.hg.1	11.75	12.89	-2.19	5.11E-06	0.0005	*RUNX1*	runt-related transcription factor 1
TC1700012185.hg.1	5.66	6.79	-2.19	0.0007	0.0135	*TNFSF12*	tumor necrosis factor (ligand) superfamily, member 12
TC0100010112.hg.1	8.17	9.29	-2.18	4.66E-06	0.0004	*SLC50A1*	solute carrier family 50 (sugar efflux transporter), member 1
TC0100007789.hg.1	9.64	10.76	-2.18	5.45E-05	0.0025	*ago-03*	argonaute RISC catalytic component 3
TC0500008479.hg.1	13.74	14.86	-2.18	0.0002	0.0061	*SNX2*	sorting nexin 2
TC1300008439.hg.1	7.12	8.25	-2.18	0.0080	0.0777	*MTIF3*	mitochondrial translational initiation factor 3
TC0700011298.hg.1	7.58	8.69	-2.17	1.04E-05	0.0007	*ZNF680*	zinc finger protein 680
TC1700008862.hg.1	6.16	7.28	-2.17	3.10E-05	0.0016	*RAB37*	RAB37, member RAS oncogene family
TC0500009061.hg.1	7.48	8.6	-2.17	0.0001	0.0047	*PCYOX1L*	prenylcysteine oxidase 1 like
TC1700010540.hg.1	11.48	12.59	-2.16	7.15E-06	0.0006	*PCGF2*	polycomb group ring finger 2
TC1000010482.hg.1	5.55	6.67	-2.16	4.89E-05	0.0023	*C10orf10*	chromosome 10 open reading frame 10
TC0500010615.hg.1	4.74	5.85	-2.16	0.0011	0.0190	*SEPP1*	selenoprotein P, plasma, 1
TC0900008793.hg.1	5.21	6.33	-2.16	0.0015	0.0237	*ZBTB34*	zinc finger and BTB domain containing 34
TC0400007853.hg.1	13.34	14.44	-2.15	1.51E-06	0.0002	*EREG*	epiregulin
TC1100012230.hg.1	9.73	10.84	-2.15	3.74E-06	0.0004	*EXPH5*	exophilin 5
TC0100011096.hg.1	7.46	8.57	-2.15	1.03E-05	0.0007	*NEK7*	NIMA-related kinase 7
TC0200009443.hg.1	6.11	7.22	-2.15	5.46E-05	0.0025	*R3HDM1*	Transcript Identified by AceView, Entrez Gene ID(s) 23518
TC2100008385.hg.1	8.56	9.66	-2.15	6.11E-05	0.0027	*SUMO3*	small ubiquitin-like modifier 3
TC0300006791.hg.1	6.54	7.64	-2.15	8.45E-05	0.0033	*KAT2B*	K(lysine) acetyltransferase 2B
TC1600008034.hg.1	7.32	8.42	-2.15	0.0003	0.0087	*NDRG4*	NDRG family member 4

TC1500009508.hg.1	4.59	5.69	-2.15	0.0008	0.0151	*NEDD4*	neural precursor cell expressed, developmentally down-regulated 4, E3 ubiquitin protein ligase
TC0800011634.hg.1	4.25	5.34	-2.14	1.02E-06	0.0001	*SNTB1*	syntrophin, beta 1 (dystrophin-associated protein A1, 59kDa, basic component 1)
TC0300014021.hg.1	10.97	12.06	-2.14	1.13E-05	0.0008	*CPOX*	coproporphyrinogen oxidase
TC0800009735.hg.1	11.3	12.4	-2.14	2.49E-05	0.0014	*MTUS1*	microtubule associated tumor suppressor 1
TC0700013468.hg.1	6.91	8.01	-2.14	4.19E-05	0.0020	*NDUFB2*	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 2, 8kDa
TC0200012713.hg.1	5.76	6.86	-2.14	0.0004	0.0095	*BCL11A*	B-cell CLL/lymphoma 11A (zinc finger protein)
TC1500009457.hg.1	7.31	8.4	-2.14	0.0013	0.0211	*MYO5A*	myosin VA
TC0500008539.hg.1	7.29	8.39	-2.14	0.0307	0.1858	*GRAMD3*	GRAM domain containing 3
TC0300012935.hg.1	7.45	8.54	-2.13	9.25E-06	0.0007	*SHOX2*	short stature homeobox 2
TSUnmapped00000398.hg.1	8.31	9.4	-2.13	1.61E-05	0.0010	*KAT6B*	K(lysine) acetyltransferase 6B
TC2000006721.hg.1	8.59	9.68	-2.13	7.07E-05	0.0030	*BTBD3*	BTB (POZ) domain containing 3
TC0900010920.hg.1	8.14	9.23	-2.13	9.29E-05	0.0035	*AAED1*	AhpC/TSA antioxidant enzyme domain containing 1
TC0800007221.hg.1	4.4	5.49	-2.13	0.0002	0.0049	*NRG1*	neuregulin 1
TC1900011233.hg.1	7.67	8.76	-2.13	0.0004	0.0101	*KLK10*	kallikrein related peptidase 10
TC0100010393.hg.1	3.33	4.41	-2.12	0.0002	0.0056	*RGS4*	regulator of G-protein signaling 4
TC0900011501.hg.1	7.34	8.42	-2.12	0.0002	0.0069	*NR6A1*	nuclear receptor subfamily 6, group A, member 1
TC1100010897.hg.1	9.81	10.9	-2.12	0.0011	0.0194	*UBE2L6*	ubiquitin-conjugating enzyme E2L 6
TC1100013223.hg.1	3.74	4.83	-2.12	0.0016	0.0244	*CARD17*	caspase recruitment domain family, member 17
TC0900011981.hg.1	4.84	5.92	-2.12	0.0111	0.0977	*NOTCH1*	notch 1
TC0X00010877.hg.1	10.31	11.39	-2.11	5.91E-06	0.0005	*FAM122B*	family with sequence similarity 122B
TC1100013059.hg.1	10.58	11.66	-2.11	1.35E-05	0.0009	*MYEOV*	myeloma overexpressed
TC1000008354.hg.1	5.76	6.83	-2.11	0.0007	0.0143	*PAPSS2*	3-phosphoadenosine 5-phosphosulfate synthase 2
TC1900008015.hg.1	4.25	5.33	-2.11	0.0008	0.0157	*SPRED3*	sprouty-related, EVH1 domain containing 3
TC0100018236.hg.1	9.49	10.56	-2.1	1.12E-05	0.0008	*ACADM*	acyl-CoA dehydrogenase, C-4 to C-12 straight chain
TC1900007290.hg.1	5.81	6.88	-2.1	9.37E-05	0.0035	*TMEM38A*	transmembrane protein 38A
TC0500013140.hg.1	7.55	8.62	-2.1	0.0015	0.0237	*AHRR; PDCD6*	aryl-hydrocarbon receptor repressor; programmed cell death 6
TC1700010996.hg.1	6.43	7.5	-2.09	2.69E-05	0.0015	*HOXB8*	homeobox B8
TC0800010285.hg.1	9.74	10.8	-2.09	0.0001	0.0043	*SLC20A2*	solute carrier family 20 (phosphate transporter), member 2
TC1400008382.hg.1	8.4	9.46	-2.09	0.0005	0.0107	*TNFAIP2*	tumor necrosis factor, alpha-induced protein 2
TC0900007518.hg.1	3.99	5.05	-2.09	0.0005	0.0110	*MAMDC2*	MAM domain containing 2
TC0800011881.hg.1	6.9	7.97	-2.09	0.0010	0.0182	*NDRG1*	N-myc downstream regulated 1
TC0700008367.hg.1	6.1	7.16	-2.09	0.0212	0.1488	*PEG10*	paternally expressed 10
TC0500007337.hg.1	10.03	11.09	-2.08	4.20E-06	0.0004	*PARP8*	poly(ADP-ribose) polymerase family member 8
TC1700012457.hg.1	7.33	8.39	-2.08	9.18E-06	0.0007	*AXIN2*	axin 2
TC1100009330.hg.1	9.37	10.43	-2.08	6.12E-05	0.0027	*UBASH3B*	ubiquitin associated and SH3 domain containing B
TC1100012583.hg.1	10.6	11.66	-2.08	8.31E-05	0.0032	*BLID; MIR100HG*	BH3-like motif containing, cell death inducer; mir-100-let-7a-2 cluster host gene
TC0900010386.hg.1	8.33	9.39	-2.08	9.33E-05	0.0035	*TMEM2*	transmembrane protein 2
TC0300011135.hg.1	7.36	8.41	-2.08	0.0001	0.0043	*VPRBP*	Vpr (HIV-1) binding protein
TC1100009026.hg.1	9.22	10.28	-2.08	0.0002	0.0066	*SIK2*	salt-inducible kinase 2

TC0X00010136.hg.1	8.44	9.49	-2.08	0.0003	0.0078	*TAF9B*	TAF9B RNA polymerase II, TATA box binding protein (TBP)-associated factor, 31kDa
TC1000008658.hg.1	5.47	6.53	-2.08	0.0006	0.0133	*PAX2*	paired box 2
TC1900011770.hg.1	6.3	7.36	-2.08	0.0010	0.0178	*C5AR1*	complement component 5a receptor 1
TC0300012339.hg.1	4.63	5.69	-2.08	0.0024	0.0334	*GATA2*	GATA binding protein 2
TC1500010162.hg.1	5.57	6.63	-2.08	0.0038	0.0459	*RASGRF1*	Ras protein-specific guanine nucleotide-releasing factor 1
TC1700006744.hg.1	6.66	7.72	-2.08	0.0184	0.1367	*NLGN2*	neuroligin 2
TC1600006593.hg.1	9.52	10.57	-2.07	1.65E-06	0.0002	*RAB26*	RAB26, member RAS oncogene family
TC0300012029.hg.1	12.67	13.72	-2.07	1.71E-06	0.0002	*NAA50*	N(alpha)-acetyltransferase 50, NatE catalytic subunit
TC0500011751.hg.1	9.74	10.79	-2.07	6.50E-06	0.0005	*FEM1C*	fem-1 homolog c (C. elegans)
TC0100013339.hg.1	7.8	8.85	-2.07	1.33E-05	0.0009	*RUNX3*	runt-related transcription factor 3
TC1000008954.hg.1	9.55	10.6	-2.07	4.49E-05	0.0022	*CASP7*	caspase 7
TC0200011171.hg.1	11.49	12.54	-2.07	0.0003	0.0072	*SH3BP4*	SH3-domain binding protein 4
TC0400007868.hg.1	7.46	8.51	-2.07	0.0028	0.0372	*PARM1*	prostate androgen-regulated mucin-like protein 1
TC0100017761.hg.1	6.49	7.55	-2.07	0.0199	0.1432	*PCNXL2*	pecanex-like 2 (Drosophila)
TC2200008799.hg.1	11.7	12.74	-2.06	2.49E-06	0.0003	*ST13*	suppression of tumorigenicity 13 (colon carcinoma) (Hsp70 interacting protein)
TC0100012222.hg.1	8.75	9.8	-2.06	3.93E-06	0.0004	*DESI2*	desumoylating isopeptidase 2
TSUnmapped00000818.hg.1	4.14	5.18	-2.06	3.02E-05	0.0016	*F10*	coagulation factor X
TC1600011517.hg.1	8.55	9.59	-2.06	6.92E-05	0.0029	*DOC2A*	double C2-like domains, alpha
TC0300012665.hg.1	6.28	7.32	-2.06	0.0035	0.0436	*XRN1*	5-3 exoribonuclease 1
TC1900010716.hg.1	7.19	8.23	-2.06	0.0104	0.0931	*ADCK4*	aarF domain containing kinase 4
TC1900007399.hg.1	6.87	7.91	-2.05	1.29E-05	0.0009	*TMEM59L*	transmembrane protein 59-like
TC1800008399.hg.1	6.76	7.8	-2.05	6.59E-05	0.0028	*B4GALT6*	UDP-Gal:betaGlcNAc beta 1,4-galactosyltransferase, polypeptide 6
TC0600010002.hg.1	6.71	7.75	-2.05	0.0008	0.0151	*SYNJ2*	synaptojanin 2
TC1600007811.hg.1	10.48	11.52	-2.05	0.0011	0.0190	*PAPD5*	PAP associated domain containing 5
TC0300007296.hg.1	6.61	7.65	-2.05	0.0226	0.1538	*SCAP*	Memczak2013 ANTISENSE, CDS, coding, INTERNAL best transcript NM_012235
TC0700012813.hg.1	5.01	6.05	-2.05	0.0331	0.1944	*DENND2A*	DENN/MADD domain containing 2A
TC0700011518.hg.1	7.7	8.72	-2.04	3.92E-06	0.0004	*GTF2IRD2; GTF2IRD2B*	GTF2I repeat domain containing 2; GTF2I repeat domain containing 2B
TC0800009927.hg.1	4.25	5.27	-2.04	0.0001	0.0037	*EBF2*	early B-cell factor 2
TC0500012909.hg.1	4.61	5.64	-2.04	0.0002	0.0068	*DRD1*	dopamine receptor D1
TC1200010284.hg.1	8.66	9.69	-2.04	0.0006	0.0129	*DENND5B*	DENN/MADD domain containing 5B
TC0400011144.hg.1	6.94	7.97	-2.04	0.0011	0.0191	*ANTXR2*	anthrax toxin receptor 2
TC1700011919.hg.1	10.88	11.9	-2.03	3.62E-06	0.0004	*CEP295NL; TIMP2*	CEP295 N-terminal like; TIMP metallopeptidase inhibitor 2
TC1000011938.hg.1	12.28	13.31	-2.03	3.37E-05	0.0017	*SHTN1*	shootin 1
TC0800010918.hg.1	6.72	7.74	-2.03	4.64E-05	0.0022	*ZNF704*	zinc finger protein 704
TC0500008835.hg.1	7.89	8.91	-2.03	0.0001	0.0048	*CXXC5*	CXXC finger protein 5
TC0400012952.hg.1	6.72	7.74	-2.03	0.0002	0.0069	*BDH2*	3-hydroxybutyrate dehydrogenase, type 2
TC1000006466.hg.1	6.58	7.6	-2.03	0.0006	0.0133	*WDR37*	WD repeat domain 37
TC1900006685.hg.1	4.44	5.46	-2.03	0.0045	0.0519	*FSD1*	fibronectin type III and SPRY domain containing 1
TC1600006537.hg.1	6.26	7.28	-2.03	0.0062	0.0656	*BAIAP3*	BAI1-associated protein 3

TC0200010416.hg.1	6.6	7.62	-2.03	0.0286	0.1783	*SPATS2L*	Transcript Identified by AceView, Entrez Gene ID(s) 26010
TC1300007774.hg.1	11	12.02	-2.02	5.06E-06	0.0005	*MBNL2*	muscleblind-like splicing regulator 2
TC0200016717.hg.1	9.34	10.36	-2.02	0.0001	0.0043	*TBC1D8*	TBC1 domain family, member 8 (with GRAM domain)
TC1200008843.hg.1	4.93	5.94	-2.02	0.0002	0.0061	*FAM216A*	family with sequence similarity 216, member A
TC2200007312.hg.1	14.42	15.44	-2.02	0.0004	0.0100	*LGALS1*	lectin, galactoside-binding, soluble, 1
TC0600007821.hg.1	5.23	6.24	-2.02	0.0014	0.0224	*LHFPL5*	lipoma HMGIC fusion partner-like 5
TC1900012025.hg.1	7.81	8.83	-2.02	0.0163	0.1257	*ZNF28*	zinc finger protein 28
TC0600011252.hg.1	4.94	5.95	-2.02	0.0328	0.1934	*ZSCAN31*	zinc finger and SCAN domain containing 31
TC1700011313.hg.1	7.9	8.91	-2.01	1.19E-05	0.0008	*USP32*	ubiquitin specific peptidase 32
TC0200015002.hg.1	6.32	7.33	-2.01	5.02E-05	0.0023	*CHN1*	chimerin 1
TC1200007887.hg.1	8.46	9.47	-2.01	7.63E-05	0.0031	*MBD6*	methyl-CpG binding domain protein 6
TC0300013104.hg.1	9.22	10.23	-2.01	0.0001	0.0037	*EIF5A2*	eukaryotic translation initiation factor 5A2
TC0800007335.hg.1	8.43	9.44	-2.01	0.0001	0.0037	*LETM2*	leucine zipper-EF-hand containing transmembrane protein 2
TC0400007495.hg.1	7.33	8.34	-2.01	0.0002	0.0053	*DCUN1D4*	DCN1, defective in cullin neddylation 1, domain containing 4
TC1100006831.hg.1	7.63	8.64	-2.01	0.0004	0.0090	*ADM*	adrenomedullin
TC1200010229.hg.1	3.72	4.73	-2.01	0.0049	0.0558	*PTHLH*	parathyroid hormone-like hormone
TC0100006681.hg.1	7.31	8.31	-2.01	0.0126	0.1064	*RNF207*	ring finger protein 207
TC0500013089.hg.1	7.36	8.37	-2.01	0.0357	0.2041	*GFPT2*	glutamine-fructose-6-phosphate transaminase 2
TC0100009142.hg.1	8.31	12	-12.91	2.21E-13	4.73E-09	*PTBP2*	polypyrimidine tract binding protein 2
TC0900012176.hg.1	8.79	7.79	2	4.10E-05	0.0020	*URM1*	ubiquitin related modifier 1
TC0100015049.hg.1	7.04	6.04	2	0.0004	0.0104	*FRRS1*	ferric-chelate reductase 1
TC0X00010837.hg.1	5.79	6.79	-2	5.73E-05	0.0026	*MBNL3*	muscleblind-like splicing regulator 3
TC1500010018.hg.1	6.16	7.16	-2	8.15E-05	0.0032	*SEMA7A*	semaphorin 7A, GPI membrane anchor (John Milton Hagen blood group)