

UNIVERSIDAD AUTÓNOMA DE MADRID

FACULTAD DE MEDICINA

**“IDENTIFICACIÓN DE NUEVOS GENES DIANA DEL FACTOR  
INDUCIBLE POR HIPOXIA HIF MEDIANTE TÉCNICAS  
COMPUTACIONALES”**

**AMAYA ORTIZ BARAHONA**

Madrid, 2010



DEPARTAMENTO DE BIOQUÍMICA  
FACULTAD DE MEDICINA  
UNIVERSIDAD AUTÓNOMA DE MADRID

**“IDENTIFICACIÓN DE NUEVOS GENES DIANA DEL FACTOR  
INDUCIBLE POR HIPOXIA HIF MEDIANTE TÉCNICAS  
COMPUTACIONALES”**

Memoria de tesis doctoral presentada por **AMAYA ORTIZ BARAHONA**, licenciada en Bioquímica, para optar al grado de doctor por la Universidad Autónoma de Madrid.

Director de tesis: **Dr. Luis del Peso Ovalle**. Profesor titular, UAM. Dpto. Bioquímica.



Luis del Peso Ovalle, profesor titular del Departamento de Bioquímica, de la Facultad de Medicina de la Universidad Autónoma de Madrid certifica que:

Amaya Ortiz Barahona, licenciada en Bioquímica por la Universidad Autónoma de Madrid, ha realizado bajo mi dirección el trabajo de investigación titulado:

*“Identificación de nuevos genes diana del factor inducible por hipoxia HIF mediante técnicas computacionales”*

Éste reúne todas las condiciones requeridas por la legislación vigente y la originalidad y calidad científica necesarias para poder ser presentado y defendido con el fin de optar al grado de Doctor.

Madrid, 24 de Mayo de 2010.

Dr. Luis del Peso Ovalle.





Esta tesis se ha realizado gracias a la concesión de la beca predoctoral (BES-2006-13438) asociada al proyecto "*Maquinaria molecular de respuesta a hipoxia. Aplicaciones terapéuticas*" (SAF2005-00180) concedido al Dr. Luis del Peso por el Ministerio de Ciencia e Innovación.





***A mis padres y a Almudena***

*Un buen científico es aquel que puede unir su imaginación con la realidad mediante el puente de la ciencia.*

A. Bernardin



## ***AGRADECIMIENTOS***



Después de cinco años de tesis, estas páginas se quedan cortas para agradecer a todos haber compartido conmigo mis muchas alegrías y penas científicas, así que os pongo un ¡¡¡GRACIAS MUY GRANDE!!!

Primeramente quiero agradecer a Luis haberme dado la oportunidad de “descubrir” la ciencia bajo su tutela. Gracias por enseñarme a “hacer ciencia” de forma estricta, sencilla y divertida, por ayudarme a desarrollar un espíritu crítico y a valorar el esfuerzo del trabajo bien hecho que hay detrás de una gráfica de 7 puntos. Ha sido un periodo de aprendizaje intelectual y de formación profesional muy importante, gracias por tu incondicional apoyo profesional y personal durante estos años.

Mi primer labo, el de Manolo en el Hospital de la Princesa, allí he pasado tantos buenos momentos científicos y personales...”la prince” tiene para mi un encanto especial. Desde el comienzo, en mi divertidísima avanzada junto a Salva y Elisa, he tenido la suerte de compartir la mayoría de estos años con todos vosotros. Sois un laboratorio estupendo a nivel científico, jefes, becarios, técnicos...todos tenéis una gran valía profesional, pero sin duda me quedo con la personal. Gracias a todos por vuestra gran ayuda durante todo estos años, que ha sido mucha, por vuestro apoyo y cariño. Gracias a mi Coqui linda “por mucho”, a mami-labo Ali por “la otra mitad de mucho”, a las dos grandes parejas: Ángel&Esther y Ainara&Bárbara, siempre con su eterna sonrisa, a Kostas por su gran espíritu científico y su dulzura, a los jefes-divertidísimos- esquiadores Mariajo y Julián, a todos los últimos fichajes que son más que estupendos, Inés, María, Dani, Raquel, Edu... etc, a “los Paco’s” que dan mucha vidilla al departamento... ¡a todos!

¡¡¡Mi cuqui-labo!!!!252-251, ¡ay mis chic@s!, con vuestra ayuda he dado el empujón final a mi tesis y habéis hecho que estos últimos años sean geniales, ¡con vosotros es un gusto llegar a trabajar cada día! Gracias a todos mis cuquis no solo por rescatarme de los aturullamientos profesionales de “¿Cómo hibrido la sonda??¿Cómo fijo las células??? ¿Cómo mato a un ratón??? ¿Cómo muto???” Sino de los personales... ¡eso sí que no tiene precio! Gracias a los de siempre, el paciente Jose, la dicharachera Benilde, la dulce Silvia, los inseparables Marisa-Diego, mis adorables-especiales-cuquitas Asun y Bea y mi Nuri, para la que tendría que hacer un apartado en exclusiva (no hace falta ni que te diga...); y a los últimos en llegar, mi mini-yo Mary, la reina de los sucesos Sara, la risueña Laurita, el más que encantador Valentí y el amoroso Alfonso. También daros las gracias a todas las personas del IIB que habéis formado parte de mi tesis haciendo muy agradable trabajar aquí, Zambrano, Josué, Irene, María, Toño, nuestros chicos de seguridad, almacén, informática, imagen...etc.

Fuera de la ciencia, tengo que agradecer a todos los que, sin saber que es la Hipoxia, se os ha cortado la respiración más de una vez aguantando estoicamente mis rollos laborales, gracias por vuestro esfuerzo Leny, Gonzalo, Fito y mi incondicional María, siempre pendiente de mi estado personal-laboral. Muy especialmente a vosotras: Serenella, que te has involucrado en mi tesis hasta el punto de coger una pipeta y mi Cristi, que ha vivido mi amor/odio por la ciencia muy de cerca desde el principio, sois ejemplo de mujeres fuertes y emprendedoras, dignas de admirar profesional y personalmente. A mis Ruthy's, por esas cenas de niñas los jueves, con sesiones estupendas de vino y rica cena aderezada con conversaciones que no queremos que acaben nunca, por esa intimidad agradable en la que me encuentro con vosotras.

Josevi, has vivido muy de cerca el año profesional más "intenso" de mi tesis, que no ha sido fácil a todos los niveles, gracias por tu buen humor y dulzura, por tu comprensión, por darme mucha "bolilla", porque crees más que yo en mi valía científica, por tu curiosidad por la ciencia, por esas interesantes conversaciones de ciencia-culinaria-vinícola, porque estando muy alejado de mi campo laboral, eres de las personas que más interés ha mostrado por conocerme hasta ese punto, porque durante este tiempo de tesis compartido me has hecho feliz!

Mis biólogos, os adoro! con vosotros empecé el camino de la ciencia hace diez años, hemos compartido "cerillismos" de: me han pisado el paper! no me sale el clonaje! no interfiero! las células se mueren o no! diosss hablar en público! pero también hemos compartido momentos profesionales muy dulces, después de mucho esfuerzo eso sí!, congresos, papers (sin importar el índice de impacto, lo importante es participar..) y las esperadas lecturas de tesis. Hemos compartido ciencia a ratitos, pero sobre todo hemos compartido una parte importante de nuestras vidas, hemos reído (y mucho!) y hemos llorado también, pero siempre juntos! sois muy importantes para mí y habéis sido un apoyo grande en mis momentos más "feos" de estos últimos años, nuestros chicos, Luisja, Coloma, Álvaro y David que nos cuidan como reinas y ...mis maris, mis niñitas, Lucy, Carol, Nagore y Lidia, tanto tanto os tengo que agradecer.., sois indispensables en mi vida y este lacito que hemos tejido durante muchos años ya se queda para siempre, os quiero muchísimo .

Mi familia, que me emocio solo de pensar en lo importante que sois para mí. Papi y Mami, vosotros habéis hecho posible esta tesis, vuestra preocupación por que fuéramos mujeres trabajadoras, independientes e intelectualmente preparadas, ha sido lo que me ha hecho llegar hasta aquí y lo que me seguirá empujando siempre. Nunca os podré agradecer

todo lo que habéis hecho por mí con vuestro amor y sacrificio incondicional, gracias por todos los valores que me habéis enseñado, sois mi referente de cómo hay que “ser de mayor”. Mi Munenita, tú has compartido todo conmigo, pequeña pero siempre con las ideas claras, con la cabeza bien alta, mi apoyo incondicional, mi flotador, la que me ha hecho pisar suelo muchas veces, mi medio “yo”... mi Mune. Y por último dedicarle la tesis a mis Belis, a los que quiero con locura, y que aunque ellos no saben de ciencia, han hecho una labor fundamental en mi educación científica, esas sesiones de hacer rosquillas y bizcocho han sido la base de cómo investigar, trabajar en equipo y cómo seguir un protocolo, para vosotros.





***RESUMEN***



La respuesta transcripcional mediada por el factor inducible por hipoxia HIF es esencial para la adaptación celular en condiciones de oxígeno limitantes. Por tanto, el conocimiento del conjunto completo de genes regulados por este factor es necesario para entender los mecanismos celulares que rigen la respuesta a hipoxia tanto en condiciones fisiológicas como patológicas. En este trabajo describimos una estrategia bioinformática, basada en la combinación de la identificación de sitios de unión de HIF (HBSs) y el meta-análisis de perfiles de expresión génica provenientes de bases de datos públicas, con la que obtenemos un listado de potenciales genes diana de HIF. La comparación de los candidatos obtenidos en este listado con los datos publicados de ensayos de ChIP-ChIP para HIF1 $\alpha$  indica un alto grado de sensibilidad (78%) y especificidad (97.7%) de la estrategia. Con el fin de validar esta estrategia llevamos a cabo ensayos de ChIP para HIF1 $\alpha$  en un conjunto de potenciales genes diana. Nuestros resultados confirmaron la fiabilidad de la estrategia a la hora de identificar sitios de unión de HIF. Como resultado de este trabajo identificamos un conjunto de 216 potenciales dianas de HIF de los que 152 no estaban previamente descritos. El análisis de ontología de este listado de estas 216 dianas reveló una nueva función biológica, hasta el momento no asociada a HIF, como es la síntesis de glucógeno. La validación funcional de cuatro genes del listado implicados en la regulación de esta vía, puso de manifiesto la importancia de esta nueva función para la adaptación celular a hipoxia mediada por HIF.

Por otro lado, aprovechamos las predicciones de esta estrategia para identificar polimorfismos en los sitios HRE predichos. Este estudio reveló que estas variaciones pueden alterar la regulación transcripcional mediada por HIF. Como muestra de ello demostramos que el SNP rs17004038, localizado en el HRE funcional del gen que codifica para el factor inhibidor de la migración de macrófagos (MIF), evita su inducción por hipoxia.

En conjunto, nuestros resultados muestran que la estrategia propuesta en este trabajo es una poderosa herramienta para identificar dianas directas de HIF, lo que permite incrementar nuestro conocimiento sobre la adaptación celular a hipoxia. Así mismo, pone de manifiesto que pueden existir variaciones inter-individuales en esta respuesta.



***ABSTRACT***



The transcriptional response driven by Hypoxia-inducible factor (HIF) is central to the adaptation to oxygen restriction. Hence, the complete identification of HIF targets is essential for understanding the cellular responses to hypoxia. Herein we describe a computational strategy based on the combination of HIF-binding sites identification and transcription profiling meta-analysis for the identification of HIF-target genes. Comparison of the resulting candidates with published HIF1a genome-wide chromatin immunoprecipitation (ChIP-ChIP) indicates a high sensitivity (78%) and specificity (97.8%). To validate this strategy, we performed HIF1a chromatin immunoprecipitation on a set of putative targets. Our results confirm the robustness of the computational strategy in predicting HIF-binding sites. The implementation of this strategy led to the identification of 216 potential HIF targets, among them, 152 were potential novel targets, whose involvement in the hypoxic response was not described. Ontology analysis of the 216 genes evidenced a new biological function dependent on HIF, glycogen synthesis. The experimental validation of four genes involved on glycogen synthesis demonstrated their importance on HIF mediated adaptation to hypoxia. In addition, mapping of described polymorphisms to the predicted HIF-binding sites identified several single-nucleotide polymorphisms (SNPs) that could alter HIF binding. As a proof of principle, we demonstrate that SNP rs17004038, mapping to a functional hypoxia response element in the macrophage migration inhibitory factor (MIF) locus, prevents induction of this gene by hypoxia. Altogether, our results show that the proposed strategy is a powerful tool for the identification of HIF direct targets that expands our knowledge of the cellular adaptation to hypoxia and provides cues on the inter-individual variation in this response.





## ***ÍNDICE***



I.ABREVIATURAS.....	1
II.INTRODUCCIÓN.....	8
II. 1. HOMEOSTASIS DEL OXÍGENO. HIPOXIA CELULAR.....	9
II.2. FACTOR INDUCIBLE POR HIPOXIA HIF.....	10
II.2.1. ESTRUCTURA Y FUNCIÓN DE HIF.....	11
II.2.2. PAPEL DE HIF $\alpha$ .....	13
II.3. REGULACIÓN DE LA ESTABILIDAD DE HIF $\alpha$ . EGLNs, SENSORES DE OXÍGENO.....	15
II.3.1. REGULACIÓN DE LAS EGLNs.....	18
II.3.2. PAPEL DE LAS EGLNs.....	19
II.4. REGULACIÓN DE LA ACTIVIDAD TRANSCRIPCIONAL DE HIF. ....	20
II.4.1. OTRAS MODIFICACIONES QUE REGULAN A HIF.....	21
II.5. TRANSCRIPTOMA DE HIF.....	22
II.5.1. COFACTORES TRANSCRIPCIONALES.....	25
II.5.2. INTERACCIÓN DE HIF $\alpha$ CON OTRAS RUTAS CELULARES.....	26
II.6. IDENTIFICACIÓN DE DIANAS DE HIF.....	26
II.6.1. TÉCNICAS COMPUTACIONALES.....	27
III.OBJETIVOS.....	29
IV.MATERIALES Y MÉTODOS.....	33
IV.1. CULTIVOS CELULARES Y TRATAMIENTO DE HIPOXIA.....	35
IV.2. GENERACIÓN DE PLÁSMIDOS.....	35
IV.3. ENSAYOS REPORTEROS.....	36
IV.4. INMUNOPRECIPITACIÓN DE CROMATINA (ChiP-Qpcr-RT).....	36
IV.5. EXTRACCIÓN DE RNA Y q-PCR-RT.....	37
IV.6. ANÁLISIS DE LOS DATOS DE LOS PERFILES DE EXPRESIÓN DE LA BASE DE DATOS <i>GEO</i> . 39	
IV.7. IDENTIFICACIÓN Y PUNTUACIÓN DE LOS SITIOS DE UNIÓN DE HIF (HBS).....	42
IV.8. CLASIFICACIÓN DE LOS GENES IDENTIFICADOS EN: DIANAS/ NO DIANAS DE HIF. ....	45
V.RESULTADOS.....	47

V.1. ESTRATEGIA BIOINFORMÁTICA PARA LA IDENTIFICACIÓN DE NUEVAS DIANAS DE HIF.	49
V.1.2. ANÁLISIS DEL PATRÓN DE REGULACIÓN DE LA EXPRESIÓN GÉNICA EN CÉLULAS SOMETIDAS A HIPOXIA.	50
V.1.3. IDENTIFICACIÓN IN SILICO DE SITIOS DE UNIÓN DE HIF EN LA TOTALIDAD DEL GENOMA.	55
V.1.4. MODELO PROBABILÍSTICO QUE INTEGRA DATOS DE EXPRESIÓN GÉNICA E IDENTIFICACIÓN DE HBSs.	59
V.2. VALIDACIÓN EXPERIMENTAL DE GENES DIANA DE HIF PREDICHOS POR NUESTRA ESTRATEGIA BIOINFORMÁTICA.	61
V.3. ESTUDIO FUNCIONAL DE LAS PREDICCIONES BIOINFORMÁTICAS.	64
V.3.1. NUEVOS GENES DIANA DE HIF IMPLICADOS EN CAMBIOS EN EL METABOLISMO DE LA GLUCOSA EN CONDICIONES DE HIPOXIA.	66
V.4. IDENTIFICACIÓN DE SNPs QUE SE LOCALIZAN EN SITIOS DE UNIÓN DE HIF.	68
VI.DISCUSIÓN	71
VII.CONCLUSIONES	81
VIII.BIBLIOGRAFÍA	85
IX.ANEXOS	95
IX.1 ANEXO I	97
IX.ANEXO II	127
IX.ANEXO III	129
X.ANEXO IV	131
IX.ANEXO V	137
IX. ANEXO VI	141
X.ARTÍCULOS RELACIONADOS	145

## ***ABREVIATURAS***



**ADM:** *Adrenomedullin*. Adrenomedulina

**AP-1:** *Transcription factor AP-1*. Factor de transcripción AP-1

**ATF-1:** *Cyclic AMP-dependent transcription factor ATF1*. Factor de transcripción dependiente de AMP cíclico.

**BHLHB-2:** *Class E basic helix-loop-helix protein 40*

**BNIP-3:** *BCL2/adenovirus E1B 19 kDa protein-interacting protein 3*.

**C3ORF28:** *E2-induced gene 5 protein*.

**CARHSP-1:** *Calcium-regulated heat stable protein 1*.

**CCNG2:** *Cyclin-G2*.

**ChIP:** *Chromatin Immunoprecipitation*. Inmunoprecipitación de cromatina.

**CITED2:** *Cbp/p300-interacting transactivator 2*. Transactivador del complejo CBP/p300.

**CNS:** *Conserved Non Coding Sequence*. Secuencias no codificantes conservadas evolutivamente.

**C-ODD:** *Carboxy-terminal Oxygen Degradation Domain*. Dominio de degradación de oxígeno carboxilo terminal.

**CREB-1:** *Cyclic AMP-responsive element-binding protein 1*. Proteína 1 de unión al Elemento de respuesta a AMP cíclico.

**Ct:** *Threshold Cycle*.

**C-TAD:** *Carboxy-terminal Transactivation Domain*. Dominio de transactivación carboxilo terminal.

**CXCR4:** *C-X-C chemokine receptor type 4*. Receptor de Quimioquina C-X-C tipo 4.

**DMOG:** *Dimethyloxallyl Glycine*.

**DMT-1:** *Natural resistance-associated macrophage protein 2*. Factor 2 de Resistencia asociado a macrófagos.

**EGLN1:** *Egl nine homolog 1*. Homólogo 1 de Egl-nine C.Elegans.

**EGLN2:** *Egl nine homolog 2*. Homologo 2 de Egl-nine C.Elegans.

**EGLN3:** *Egl nine homolog 3*. Homólogo 3 de Egl-nine C.Elegans.

**EMSA:** *Electrophoretic mobility shift assay*.

**EPO:** *Erythropoietin*. Eritropoyetina.

**FDR:** *False Discovery Rate*.

**FIH:** *Hypoxia-inducible factor 1-alpha inhibitor*. Inhibidor de HIF.

**GBE-1:** *1,4-alpha-glucan-branching enzyme*. Enzima ramificadora 1-4 de glucanos.

**GLUT-1:** *Glucose transporter type 1*. Transportador de glucose tipo 1.

**GYS-1:** *Muscle glycogen synthase*. Sintasa de glucógeno muscular.

**HBS:** *HIF binding site*. Sitio de unión de HIF.

## Abreviaturas

**HIF:** *Hypoxia Inducible Factor*. Factor Inducible por Hipoxia.

**HIVEP1:** *Zinc finger protein 40*

**HK-1:** *Hexokinase-1*. Hexokinasa-1.

**HNF4:** *Hepatocyte nuclear factor 4-alpha*. Factor nuclear 4 $\alpha$  de hepatocitos.

**HO-1:** *Heme oxygenase 1*. Hemo oxigenasa 1.

**HOXA4:** *Homeobox protein Hox-A4*

**HRE:** *Hypoxia responsive element*. Elemento de respuesta a hipoxia.

**Hsc-70:** *Heat shock cognate 70 kDa protein*

**Hsp70:** *Heat shock 70 kDa protein*.

**ID:** *Inhibitory domain*. Dominio de inhibición

**IDH-2:** *Isocitrate dehydrogenase [NADP], mitochondrial*. Isocitrato deshidrogenasa mitocondrial.

**IGFBP-3:** *Insulin-like growth factor-binding protein 3*. Proteína de unión 3 de IGF.

**IL-8:** *Interleukin-8*. Interleukina 8.

**iNOS:** *Nitric oxide synthase, inducible*. Sintasa de oxido nítrico inducible.

**IRF1:** *Interferon regulatory factor 1*

**IRS-4:** *Insulin receptor substrate 4*.

**LDHA:** *L-lactate dehydrogenase A chain*. Cadena A de la lactato deshidrogenasa.

**LTBP1:** *Latent-transforming growth factor beta-binding protein 1*

**MAFF:** *Transcription factor MafF*.

**MIF:** *Macrophage migration inhibitory factor*. Factor inhibidor de migración de macrófagos.

**MKRN-1:** *E3 ubiquitin-protein ligase makorin-1*.

**MMP1:** *Interstitial collagenase*. Colagenasa intersticial.

**MYC:** *Myc proto-oncogene protein*. Proteína pro-oncogénica Myc.

**NDRG1:** *N-myc downstream-regulated gene 1 protein*.

**NFkB NEMO:** *NF-kappa-B essential modulator*.

**NFkB:** *Nuclear factor NF-kappa-B*. Factor nuclear NF-Kappa-B

**N-ODD:** *Amino-terminal Oxygen Degradation Domain*. Dominio de degradación de oxígeno amino terminal.

**NOTCH:** *Neurogenic locus notch homolog protein*.

**N-TAD:** *Amino-terminal Transactivation Domain*. Dominio de transactivación amino terminal

**ODD:** *Oxygen Degradation Domain*. Dominio de degradación dependiente de oxígeno.

**P4H-TM:** *Transmembrane prolyl 4-hydroxylase*. Prolil hidroxilasa 4 transmembrana.



**PDGF:** *Platelet-derived growth factor*. Factor de crecimiento derivado de plaquetas.

**PKC- $\alpha$ :** *3-phosphoinositide-dependent protein kinase 1*.

**PKC- $\alpha$ :** *3-phosphoinositide-dependent protein kinase 1*. Proteína quinasa 1 dependiente de 3-fosfoinositol.

**PKC- $\beta$ :** *Phosphoglycerate kinase 1*. Fosfoglicerato quinasa 1.

**PKC- $\gamma$ :** *Prolyl hydroxylase 1*. Hidroxilasa de prolina 1.

**PKC- $\delta$ :** *Prolyl hydroxylase 2*. Hidroxilasa de prolina 2.

**PKC- $\epsilon$ :** *Prolyl hydroxylase 3*. Hidroxilasa de prolina 3.

**PKC- $\zeta$ :** *p21-activated protein kinase-interacting protein 1*.

**PPP1R3C:** *Protein phosphatase 1 regulatory subunit 3C*.

**PWM:** *Position Weight Matrix*.

**qPCR-RT:** *Quantitative PCR real time*

**REF-1:** *Transcriptional coactivator Aly/REF*. Coactivador transcripcional Aly/REF.

**RPLP0:** *60S acidic ribosomal protein P0*.

**Siah1 $\alpha$ /2:** *E3 ubiquitin-protein ligase SIAH1/2*. E3 proteína ubiquitín ligasa SIAH1/2.

**SMAD3:** *Mothers against decapentaplegic homolog 3*.

**SRC1:** *Nuclear receptor coactivator 1*. Coactivador 1 de receptor nuclear.

**STAT-3:** *Signal transducer and activator of transcription 3.1*.

**STT3A:** *Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3A*.

**TGF $\alpha$ / $\beta$ :** *Transforming growth factor alpha/beta*.

**TIF2:** *Nuclear receptor coactivator 2*. Coactivador 2 de receptor nuclear.

**TPD-52:** *Tumor protein D52*.

**UGP-2:** *UTP--glucose-1-phosphate uridylyltransferase*. Uridiltransferasa de UTP-glucosa-1-fosfato.

**VEGFA:** *Vascular endothelial growth factor A*. Factor de crecimiento A del endotelio vascular.

**VHL:** *Von Hippel-Lindau disease tumor suppressor*. Supresor tumoral Von Hippel-Lindau.

**ZNF395:** *Zinc finger protein 395*.



## ***INTRODUCCIÓN***



## II. 1. HOMEOSTASIS DEL OXÍGENO. HIPOXIA CELULAR.

El oxígeno es la molécula más abundante en nuestro planeta y fue un factor determinante para la evolución de la vida en la tierra hace 2.000 millones de años. Así, el paso de una atmósfera reductora a la actual atmósfera (21% de oxígeno) permitió el desarrollo de mecanismos de obtención de energía altamente eficientes, lo que llevó a una explosión y diversificación de formas de vida aerobias (Webster 2007).

El éxito de esta molécula radica en su papel como aceptor de electrones en el metabolismo aerobio de eucariotas; el oxígeno es utilizado por estos organismos durante el catabolismo de nutrientes para la generación de compuestos de alta energía, siendo mayor su consumo cuanto mayor tamaño y actividad metabólica tiene el organismo. Sin embargo, el oxígeno es una molécula altamente reactiva, pudiendo generar especies con electrones desapareados que producen daños en las macromoléculas celulares como DNA, RNA y proteínas e incluso inducir muerte celular. Por este motivo los organismos han desarrollado sistemas que permiten regular la homeostasis de este compuesto, manteniendo un fino equilibrio entre el aporte y consumo del mismo (Semenza 2004). De este modo, a lo largo de la evolución, se han desarrollado sistemas tan sofisticados como los sistemas circulatorio y respiratorio de mamíferos, que, de forma coordinada y regulados por mecanismos neuronales y hormonales, responden a variaciones puntuales en el aporte del mismo. A nivel sistémico, una de estas respuestas agudas a la falta de oxígeno se inicia cuando el cuerpo carotideo detecta un descenso de la concentración de oxígeno en sangre; sus células glómicas sufren un cambio de polaridad de membrana, lo que promueve la liberación de neurotransmisores (dopamina) de manera dependiente de calcio, esto induce un incremento de la frecuencia cardíaca y respiratoria, restableciéndose los niveles normales de oxígeno en sangre y por tanto el aporte del mismo a los diferentes tejidos (Lopez-Barneo, Ortega-Saenz et al. 2009).

Cuando esta disminución de oxígeno se prolonga en el tiempo, hipoxia crónica, se activan mecanismos celulares que implican cambios en la expresión génica. Uno de los primeros genes relacionado con la respuesta celular a hipoxia es el gen de la eritropoyetina (EPO), esta proteína estimula la proliferación y supervivencia de los progenitores de los eritrocitos de la sangre. Estudios iniciales demostraron que la cantidad de esta proteína aumentaba en condiciones de hipoxia y el análisis de sus elementos en *cis* llevó a la identificación del sitio de unión del factor inducible por hipoxia HIF (HRE) y por tanto, al clonaje y caracterización de este factor transcripcional como regulador de EPO en condiciones de hipoxia (Semenza and Wang 1992). Aunque la expresión de EPO está restringida a tejido

hepático, HIF se expresa de forma ubicua en todos los tejidos actuando, a nivel celular, como el regulador central de la respuesta a hipoxia. Esta respuesta conlleva la expresión de múltiples genes implicados en metabolismo anaerobio, angiogénesis, eritropoyesis, etc. que permiten la adaptación celular a la bajada de oxígeno (fig1).

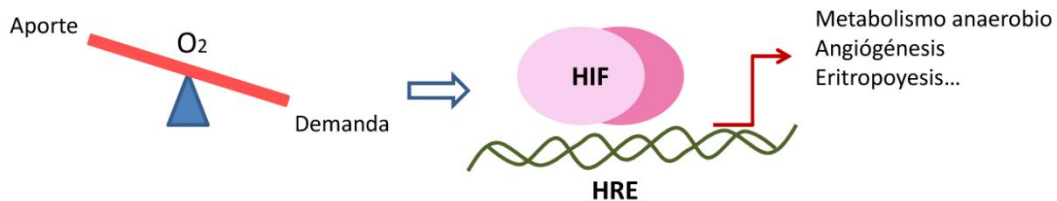


Fig 1: HIF media la transcripción de genes implicados en funciones que permiten la adaptación celular a hipoxia.

Además de HIF, existen otros factores cuya actividad transcripcional se ve incrementada en condiciones de hipoxia, como son algunos miembros de la familia de NFkB (Perkins and Gilmore 2006) y AP1 (Laderoute 2005), pero los mecanismos de activación de estos factores así como su importancia en la respuesta a hipoxia están poco claros. Por ello se considera a HIF como el principal factor que media la adaptación a hipoxia.

Todas las medidas de control descritas permiten el perfecto mantenimiento del aporte de oxígeno a nivel celular y tisular, de modo que solo en situaciones fisiológicas puntuales, como el desarrollo embrionario, o patológicas, como tumores sólidos o procesos inflamatorios, se darían situaciones de desbalance entre el aporte y consumo de oxígeno.

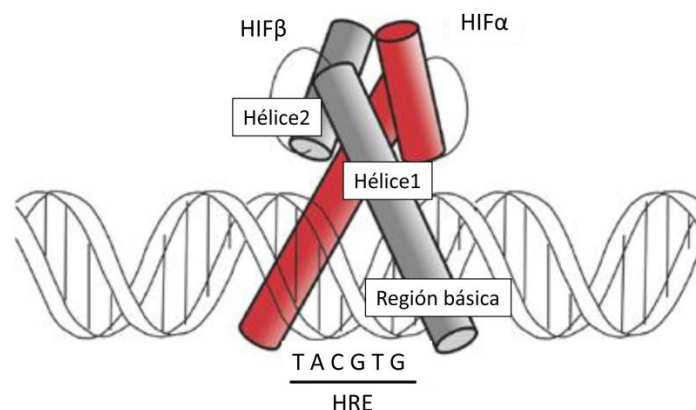
## II.2. FACTOR INDUCIBLE POR HIPOXIA HIF.

El factor inducible por hipoxia se ha identificado en múltiples especies de metazoos desde *C. Elegans* hasta *H. Sapiens*, lo que sugiere que la aparición de HIF fue un mecanismo adaptativo esencial en su proceso evolutivo. La importancia de este factor en organismos superiores se pone de manifiesto cuando se delecciona este gen en homocigosis, en ratones provoca letalidad embrionaria debido a fallos en el desarrollo placentario y el sistema circulatorio (Iyer, Leung et al. 1998).

### II.2.1. ESTRUCTURA Y FUNCIÓN DE HIF.

HIF es una proteína heterodimérica perteneciente a la familia de los factores de transcripción bHLH-PAS por presentar un dominio bHLH (*basic hélix loop hélix*) y un dominio tipo PAS (homólogo a los factores *Per*, *ARNT* y *Sim*), ambos dominios están implicados en heterodimerización con otros miembros de la familia y en unión a DNA (Iyer, Leung et al. 1998) (Wiesener, Turley et al. 1998) (Fig2). Este factor está compuesto por una subunidad HIF $\alpha$ , cuya estabilidad depende de la tensión de oxígeno celular, y una subunidad HIF $\beta$  (ARNT, homología con la proteína *Aryl Hydrocarbon Tanslocator Factor*), que se expresa de forma constitutiva (Hoffman, Reyes et al. 1991). HIF $\beta$ , además, puede dimerizar con otros factores de transcripción de la familia bHLH implicados en otras funciones como la regulación de los ciclos circadianos, crecimiento y señalización celular, etc. (Yang, Guo et al. 2010).

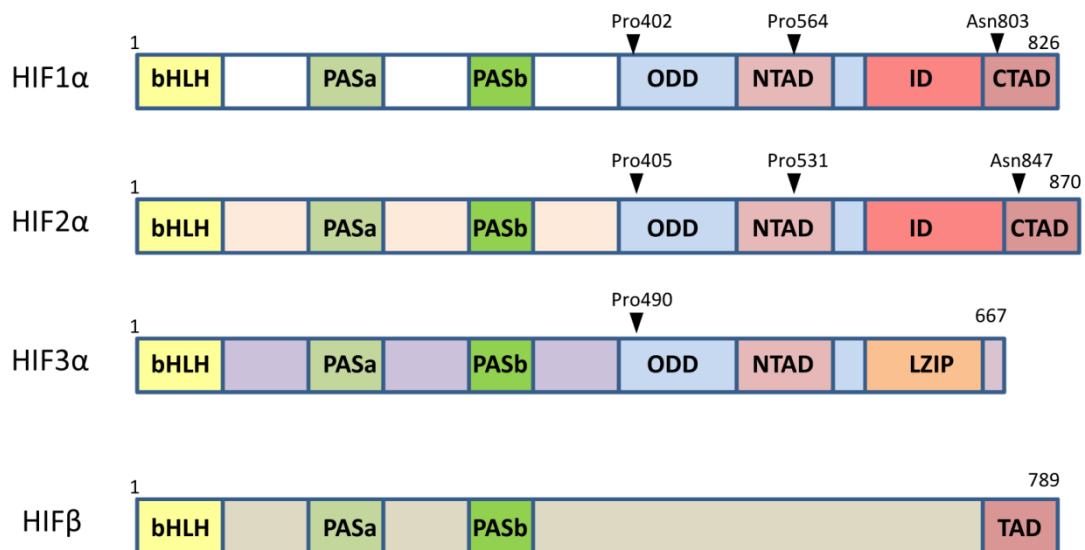
Ambas subunidades tienen una elevada similitud estructural ya que las dos presentan en su extremo amino los dominios bHLH y PAS, y en su extremo carboxilo terminal, regiones de activación transcripcional que les permiten interactuar con coactivadores. Estudios bioquímicos iniciales pusieron de manifiesto que es la subunidad  $\alpha$  la que responde a la falta de oxígeno, concretamente se observó que la vida media de esta subunidad estaba controlada por oxígeno mientras que la expresión de la subunidad  $\beta$  era insensible a la concentración del mismo (Salceda and Caro 1997; Huang, Gu et al. 1998). Posteriormente, se vio que solo la subunidad  $\alpha$  contenía un dominio implicado en la degradación dependiente de oxígeno (ODD) (fig 3) (Huang, Gu et al. 1998).



**Fig 2: Heterodimerización de HIF y unión a DNA.** El heterodímero se une al DNA por sus dominios bHLH, los residuos básicos cercanos al extremo Nterminal permiten la interacción con los nucleótidos del sitio de unión de HIF (HRE). Adaptado de (Lisy and Peet 2008).

## Introducción

Existen tres isoformas de la subunidad HIF $\alpha$ , HIF1 $\alpha$ , HIF2 $\alpha$  y HIF3 $\alpha$ , que son codificadas por tres genes independientes (fig3). Todas ellas presentan un dominio de regulación dependiente de oxígeno (ODD) (ver apartado II.3), éste contiene residuos de prolina (Pro 402/Pro564 en HIF1 $\alpha$ , Pro405/Pro531 en HIF2 $\alpha$  y Pro 490 en HIF3 $\alpha$ ) que determinan su estabilidad según la tensión de oxígeno, y al menos un dominio de transactivación, uno en HIF3 $\alpha$  y dos en el caso de HIF1 $\alpha$  y HIF2 $\alpha$ . Los dos dominios de transactivación de 1 $\alpha$  y 2 $\alpha$ , amino terminal (N-TAD amino ácidos 531–575) y carboxilo terminal (C-TAD amino ácidos 786–826) (Jiang, Zheng et al. 1997), están separados por un dominio de inhibición (O'Rourke, Tian et al. 1999). Ambos dominios de transactivación no presentan a penas similitud en su secuencia de aminoácidos, lo que puede ser indicativo de funciones divergentes (Dayan, Roux et al. 2006). Cada uno de ellos presenta un residuo diferente susceptible a la regulación por oxígeno (Prolina 564 y Aspático 803 respectivamente para HIF1 $\alpha$ ) (fig3), lo que se ha sugerido como un mecanismo dual de regulación de los genes diana de HIF en función de la tensión de oxígeno (Dayan, Roux et al. 2006).



**Fig 3: Dominios estructurales de HIF.** Las tres isoformas de HIF $\alpha$  y HIF $\beta$  presentan dominios de unión a DNA bHLH y heterodimerización PAS. HIF1 $\alpha$  y HIF2 $\alpha$  presentan dos dominios de transactivación CTAD y NTAD separados por un dominio de inhibición ID, sin embargo HIF3 $\alpha$  y HIF $\beta$  solo presentan un dominio de transactivación. Las tres HIF $\alpha$  presentan un dominio de degradación dependiente de oxígeno ODD. Se muestra también el dominio de cremallera de leucina de HIF3 $\alpha$  (LZIP). Los residuos susceptibles de prolina y aspartato marcados median la sensibilidad de HIF a oxígeno.



Como puede observarse en la figura 3 las subunidades 1 $\alpha$  y 2 $\alpha$  presentan un mayor grado de semejanza estructural, lo que se refleja en un mayor solapamiento funcional. La subunidad 3 $\alpha$  es la peor caracterizada, ya que existen múltiples isoformas resultado de procesamientos alternativos de su mensajero, de todas ellas, se ha implicado a la isoforma IPAS en inhibición de la transcripción mediada por HIF1 $\alpha$ , a la que se une impidiendo su dimerización con HIF $\beta$ . Estudios de sobreexpresión de HIF3 $\alpha$  en células de hepatoma reflejan el posible papel de esta subunidad en la disminución del crecimiento tumoral por inhibición de la angiogénesis (Makino, Cao et al. 2001).

El heterodímero HIF $\alpha$ /HIF $\beta$  se une a las secuencias RCGTG de los elementos de respuesta a hipoxia (HREs) presentes en sus genes diana, esta regulación transcripcional permite la adaptación celular a la hipoxia. La actividad transcripcional de HIF presenta diferentes matices dependiendo de la subunidad  $\alpha$  que conforme el complejo transcripcional (ver apartado II.4). Así, existe una regulación génica diferencial, que puede venir dada, en algunos casos, por las diferencias de expresión de HIF1 $\alpha$  y HIF2 $\alpha$  en los tejidos, mientras HIF1 $\alpha$  se expresa de forma ubicua 2 $\alpha$  solo lo hace en unos pocos tejidos (Gordan, Bertout et al. 2007); o bien su interacción con diferentes cofactores transcripcionales (Hu, Sataur et al. 2007) (ver apartado II.5.1).

## II.2.2. PAPEL DE HIF $\alpha$ .

### *II.2.2.1. Modelos genéticos experimentales.*

En los últimos años se han llevado a cabo numerosos estudios que pretenden determinar la importancia biológica de cada uno de los componentes de la ruta de HIF. Así se ha descrito la implicación de HIF1 $\alpha$  en desarrollo embrionario, su deficiencia en ratones provoca letalidad embrionaria debido a alteraciones en el sistema circulatorio como malformaciones cardíacas y regresión vascular (Iyer, Kotch et al. 1998) y alteraciones en el desarrollo neuronal (Ryan, Lo et al. 1998).

El defecto en HIF2 $\alpha$  también se ha asociado con letalidad embrionaria (Peng, Zhang et al. 2000), pero puesto que la muerte se da en etapas más tardías, se han podido estudiar los efectos provocados en los diferentes órganos. Debido a que HIF2 $\alpha$  se expresa solo en unos pocos tejidos como endotelio, hígado, riñón pulmón y cerebro (Tian, McKnight et al. 1997) y que el knockout para HIF2 $\alpha$  parece variar en gran medida según el fondo genético, se han descrito diferentes fenotipos. Así se ha visto que ratones deficientes en HIF2 $\alpha$  presentan

bradicardia como consecuencia de un déficit en la producción de catecolaminas (Tian, Hammer et al. 1998) suponiendo la muerte a mitad del periodo gestacional; deficiencias en la remodelación vascular primaria (Peng, Zhang et al. 2000) y alteraciones respiratorias resultado de una alteración en la producción de surfactante pulmonar (Compernelle, Brusselmans et al. 2002). Al igual que sus homólogos, HIF $\beta$  también es letal embrionario, ratones HIF $\beta$ <sup>-/-</sup> presentan importantes alteraciones vasculares como defectos de angiogénesis y hemorragias placentarias (Maltepe, 1997 #504), (Kozak, Abbott et al. 1997)

Estudios en ratones heterocigotos de HIF permitieron el estudio de los efectos de tratamientos de hipoxia crónica e intermitente, así se vio que ratones HIF1 $\alpha$ <sup>+/-</sup> desarrollaban policitemia, hipertensión pulmonar e hipertrofia ventricular (Yu, Shimoda et al. 1999). Por otro lado, la generación de ratones knockout condicionales ha permitido el análisis, en ratones adultos, de la implicación de HIF en diferentes funciones biológicas en diversos órganos. Así se ha visto que la pérdida de HIF1 $\alpha$  en células de la línea mieloide provoca alteraciones de la respuesta inmune (Cramer, Yamanishi et al. 2003) (Peyssonnaud, Datta et al. 2005); en músculo esquelético provoca daños fibrilares tras periodos de ejercicio intenso debido a un aumento de la actividad mitocondrial y una reducción de la producción de lactato (Mason, Howlett et al. 2004); en corazón y sistema vascular conlleva una disfunción contráctil (Huang, Hickey et al. 2004) y una alteración de los parámetros del endotelio vascular (Tang, Wang et al. 2004); y en tejido neuronal se ha visto que la pérdida de HIF1 $\alpha$  previene de la muerte celular asociada a procesos isquémicos (Helton, Cui et al. 2005). Por su parte, HIF2 $\alpha$  ha sido implicado en la generación de eritrocitos, su defecto produce anemias por deficiencia de expresión de EPO (Gruber, Hu et al. 2007).

Todas estas aproximaciones han puesto de manifiesto la importancia de HIF no solo en adaptación a hipoxia, sino también en la biología de cada tipo celular.

### *II.2.2.1. Implicación de HIF en patologías.*

Como hemos mencionado anteriormente, HIF media la adaptación celular a la bajada de oxígeno promoviendo la transcripción de genes que permiten restablecer el aporte de oxígeno y mantener los niveles de ATP celulares. Sin embargo, también está implicado en el desarrollo de procesos patológicos pudiendo actuar como potenciador o inhibidor de los mismos. En procesos tumorales, la expresión de HIF se ve aumentada en los núcleos hipóxicos formados durante el crecimiento tumoral (Semenza 2003), en ellos HIF promueve la expresión de VEGF, PDGF, TGF $\alpha/\beta$ , CXCR4, SDF-1 y MMP1, que favorecen la generación de nuevos vasos,

la progresión tumoral y metástasis (Kaelin 2007). Por otro lado, HIF tiene un papel protector en enfermedades cardiovasculares que cursan con hipoxia, trabajos recientes han revelado que los efectos citoprotectores están basados al menos en parte, en la reprogramación metabólica (Aragones, Schneider et al. 2008). Esta reprogramación implica una reducción del metabolismo oxidativo, que HIF promueve disminuyendo la entrada de metabolitos en el ciclo de Krebs, aumentando la glucólisis, induciendo autofagia mitocondrial y la reducción de su biogénesis, etc (Zhang, Bosch-Marce et al. 2008). Así mismo se le ha implicado a HIF en procesos de preconditionamiento hipóxico, este fenómeno se ha descrito en varios tejidos como corazón, cerebro y riñón, y supone un mecanismo protector en procesos de isquemia. Cuando un órgano es sometido a un tiempo corto de hipoxia, previo a un periodo isquémico, se pone de manifiesto una protección celular que no se observa en el caso del tejido no pre-tratado. Así, ensayos de isquemia-reperfusión llevados a cabo con ratones salvajes y HIF1 $\alpha$ +/- han puesto de manifiesto la importancia de HIF1 $\alpha$  en la protección cardiaca durante estos procesos (Cai, Zhong et al. 2008)

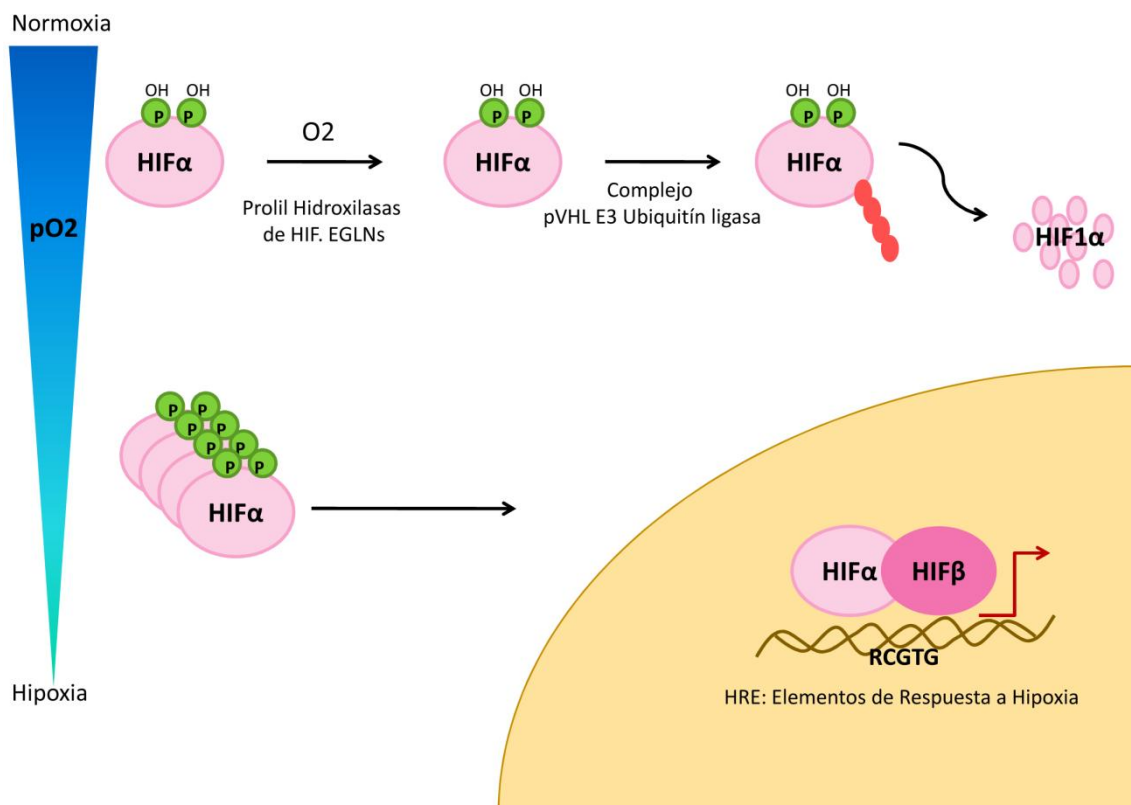
Puesto que HIF juega un importante papel en procesos patológicos que cursan con hipoxia, el conocimiento global de la respuesta transcripcional mediada por este factor puede contribuir al desarrollo de nuevas estrategias terapéuticas dirigidas al tratamiento de estas patologías.

### **II.3. REGULACIÓN DE LA ESTABILIDAD DE HIF $\alpha$ . EGLNs, SENSORES DE OXÍGENO.**

Se han propuesto numerosos mecanismos implicados en detección de la variación de la pO<sub>2</sub> celular (Semenza 1999), sin embargo, el modelo establecido se basa en la degradación de HIF en normoxia dependiente de la actividad de unas prolil hidroxilasas, EGLNs, altamente sensibles a las variaciones de pO<sub>2</sub>. En condiciones de normoxia, HIF $\alpha$  es hidroxilado por las EGLNs en dos residuos de prolina, situados en el dominio de degradación dependiente de oxígeno (ODD, N-ODD y C-ODD). La hidroxilación en al menos uno de estos residuos determina la estabilidad de HIF $\alpha$ , ya que permite su reconocimiento por parte de la proteína VHL (Von Hippel Lindau) que forma parte de un complejo E3-ubiquitín ligasa que marca a HIF $\alpha$  para su degradación vía proteosoma (Maxwell, Wiesener et al. 1999; Ivan, Kondo et al. 2001; Jaakkola, Mole et al. 2001) (fig 4).

Las EGLNs pertenecen a una familia de dioxigenasas que utilizan oxígeno y 2-Oxoglutarato como co-sustratos y hierro y ascorbato como cofactores, para catalizar la

reacción de hidroxilación. Las enzimas de esta familia son homologas a la hidroxilasa de HIF1 $\alpha$  en *C.Elegans*, EGL-9 (*Egg lying nine*) (Epstein, Gleadle et al. 2001) (Bruick and McKnight 2001) (Freeman, Hasbani et al. 2003), y su característica más importante es que tienen una Km para el oxígeno muy elevada, en torno a 200 $\mu$ M, de modo que pequeñas variaciones de la pO<sub>2</sub> en la célula provocan la disminución drástica de su actividad. Como consecuencia de ello, en hipoxia, HIF no es hidroxilado, escapa a la degradación por VHL (Ehrismann, Flashman et al. 2007) y promueve la regulación transcripcional que media la adaptación a la bajada de la pO<sub>2</sub>. Por esta característica, las EGLNs son denominadas como los sensores de oxígeno celulares (fig4).



**Fig 4: Hipoxia celular. Ruta de HIF.** En normoxia HIF $\alpha$  es hidroxilado por las HIF prolil hidroxilasas, EGLNs, en dos residuos de prolina, esto permite su reconocimiento por la proteína VHL que forma parte de un complejo E3 ubiquitín ligasa que lo marca para su degradación vía proteosoma. Cuando la pO<sub>2</sub> disminuye, las prolil hidroxilasas ven comprometida su actividad, HIF $\alpha$  se estabiliza, pasa al núcleo donde dimeriza con HIF $\beta$  y se une a los elementos de respuesta a hipoxia de sus genes diana modulando su transcripción.

Existen tres isoformas de EGLNs codificadas por genes independientes, EGLN1, EGLN2 y EGLN3 (también llamadas PHD2, PHD1 y PHD3) (Taylor 2001) y se ha propuesto una cuarta P4H-TM (Koivunen, Tiainen et al. 2007) de la que no existen aún datos experimentales concluyentes en cuanto a su función como regulador de HIF. Las tres hidroxilasas de HIF bien caracterizadas presentan diferente distribución celular (Metzen, Berchner-Pfannschmidt et al. 2003) y tisular (Lieb, Menzies et al. 2002) (tabla 1), lo que puede ser indicativo de funciones no solapantes.

Tipo de EGLN	Localización subcelular	Distribución tisular
EGLN1	Citoplasmática fundamentalmente	Ubicua
EGLN2	Nuclear fundamentalmente	Muy representada en testículo
EGLN3	Nuclear y citoplásmica	Muy representada en corazón

Tabla 1: Distribución celular y tisular de las diferentes isoformas de EGLN

Estas enzimas hidroxilan a HIF en residuos de prolina conservados dentro del motivo LXXLAP (donde X es cualquier aminoácido) (Bruick and McKnight 2001), sin embargo, cada una de las isoformas presenta especificidad por alguno de los residuos de prolina (Landazuri, Varavega et al. 2006). Todas ellas hidroxilan de forma más activa la prolina presente en el C-ODD y concretamente EGLN3 es incapaz de hidroxilar la prolina del N-ODD (Hirsila, Koivunen et al. 2003) (fig5).

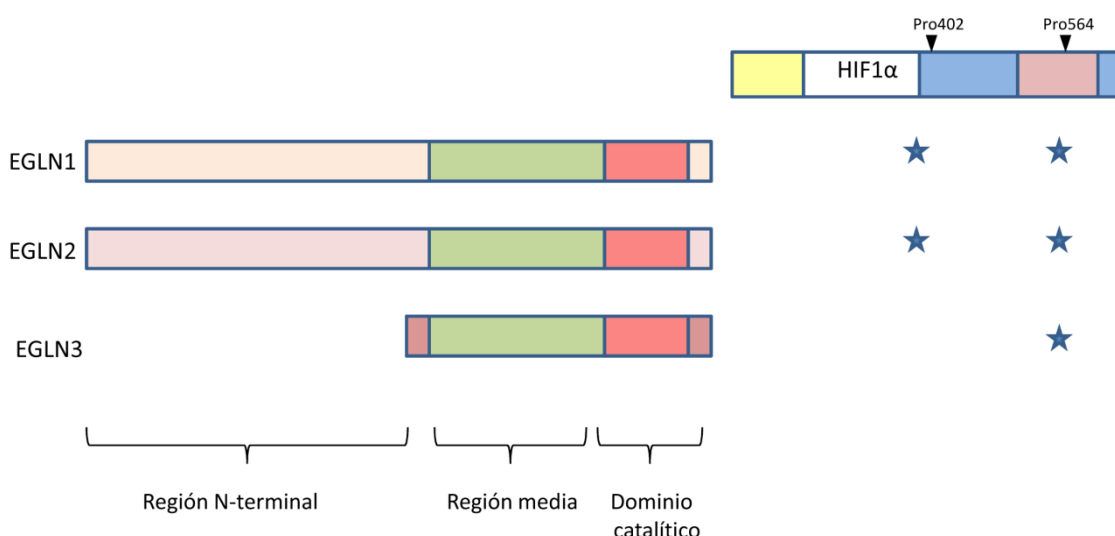


Fig 5: Especificidad de residuo para cada isoforma de EGLN. Se indica la especificidad de cada EGLN por los residuos de prolina, P402 y P564, de la isoforma HIF1α.

En cultivos celulares las tres EGLNs contribuyen a la regulación de HIF1 $\alpha$  y HIF2 $\alpha$ , pero existen preferencias. EGLN2 y EGLN3 son más activas para hidroxilar a HIF2 $\alpha$ , mientras que EGLN1 hidroxila más eficientemente a HIF1 $\alpha$  (Berra, Benizri et al. 2003). Las diferencias en cuanto a la abundancia relativa de cada isoforma en diferentes tipos celulares podrían explicar estas “preferencias”, sin embargo éstas se mantienen aún cuando los niveles de estas enzimas son similares. Estas preferencias son tan marcadas que, incluso cuando se delecciona EGLN1, EGLN3 es incapaz de desestabilizar HIF1 $\alpha$  por completo (Berra, Benizri et al. 2003; Appelhoff, Tian et al. 2004).

### II.3.1. REGULACIÓN DE LAS EGLNs.

Las HIF prolin hidroxilasas, son enzimas unidireccionales, la reacción de hidroxilación que catalizan es irreversible, de modo que su actividad enzimática viene condicionada en primer lugar por la presencia de los cofactores y cosustratos necesarios, y en segundo lugar por la cantidad de enzima disponible.

La cantidad de estas enzimas, como hemos mencionado antes (tabla 1), varía de forma tejido específica. A pesar de que la EGLN1 es la isoforma más representada en la mayoría de los tejidos, en testículos aparece como isoforma mayoritaria EGLN2 y en corazón EGLN3 (Willam, Maxwell et al. 2006).

En hipoxia la cantidad de EGLN1 y EGLN3 se ve aumentada (Epstein, Gleadle et al. 2001) y esto es debido a que son genes diana de HIF (Berra, Richard et al. 2001) (del Peso, Castellanos et al. 2003), presentan elementos de respuesta a hipoxia en secuencias reguladoras (Pescador, Cuevas et al. 2005). Por tanto, en hipoxia se produce un proceso de retroalimentación negativa en el que HIF se estabiliza ante la bajada de oxígeno y promueve la transcripción de estas dos EGLNs (del Peso, Castellanos et al. 2003), el aumento de la cantidad de estas enzimas permite una rápida degradación de HIF $\alpha$  al restablecerse la pO<sub>2</sub> celular normal (Metzen, Stiehl et al. 2005).

La vida media de estas enzimas también condiciona la cantidad de enzima disponible, y ésta viene dada, en el caso de EGLN2 y 3, por su unión con la proteína Siah1 $\alpha$ /2, que promueve su poliubiquitinación y degradación vía proteosoma; curiosamente Siah1 $\alpha$ /2 ve aumentada su expresión en hipoxia (Nakayama, Frew et al. 2004), lo que favorecería la estabilización de HIF. EGLN1 no es susceptible a la acción de este enzima y es una peptidil cis-trans isomerasa la que juega un papel en regular su estabilidad (Barth, Nesper et al. 2007).

## II.3.2. PAPEL DE LAS EGLNs

### *II.3.2.1. Modelos genéticos experimentales.*

El estudio de modelos animales deficientes en cada una de las EglNs ha permitido identificar las diferentes funciones biológicas en las que están implicados estos enzimas. Así, mientras que los ratones deficientes en EGLN2 y 3 son viables, los homocigotos para EGLN1 mueren en periodo gestacional debido a fallos en el desarrollo cardíaco y en la irrigación placentaria (Takeda, Ho et al. 2006). Sin embargo, la generación de un knockout condicional para este gen permitió detectar un aumento de la densidad de la vasculatura en la generalidad de los tejidos (Takeda, Aguila et al. 2008). El modelo de ratones EGLN2<sup>-/-</sup> reveló una protección frente al daño muscular inducido por isquemia (Aragones, Schneider et al. 2008), mientras que el modelo de EGLN3<sup>-/-</sup> pone de manifiesto la importancia de esta proteína en el sistema simpoadrenal (Bishop, Gallagher et al. 2008).

### *II.3.2.2. Implicación de las EGLNs en patologías.*

Numerosas patologías como anemia, alteraciones cardíacas, derrames cerebrales, etc, están ligados a un desajuste en la oxigenación de los tejidos. A través de la estabilización de HIF se promueve la producción de eritrocitos, vasodilatación, producción de nuevos vasos y supervivencia celular, todo ello supone un freno a estas patologías. Así, la estabilización de HIF mediante el uso de inhibidores químicos de las EGLNs, ha sido objeto de estudio en numerosos ensayos preclínicos orientados al tratamiento de ciertas patologías como anemia (Hsieh, Linde et al. 2007), y a promover efectos protectores en infartos de miocardio (Shohet and Garcia 2007), derrames cerebrales (Ratan, Siddiq et al. 2007) e isquemias renales (Bernhardt, Campean et al. 2006). Sin embargo, el tratamiento crónico con este tipo de agentes puede provocar múltiples alteraciones. Así, la estabilización de HIF mantenida en el tiempo, puede provocar policitemias, como es el caso de individuos que presentan una mutación hereditaria en EGLN1 que reduce su actividad enzimática (Percy, Zhao et al. 2006).

La identificación de las EGLNs como los reguladores principales de HIF, y por tanto sensores de oxígeno celular, ha abierto un nuevo campo de estudio orientado al desarrollo de aplicaciones farmacológicas para el tratamiento de enfermedades asociadas con la hipoxia.

#### **II.4. REGULACIÓN DE LA ACTIVIDAD TRANSCRIPCIONAL DE HIF.**

La activación de la transcripción mediada por HIF en hipoxia requiere varios eventos como es la síntesis de HIF $\alpha$ , su estabilización, el paso al núcleo, su dimerización con HIF $\beta$ , la unión a DNA y el reclutamiento de coactivadores y maquinaria transcripcional. Sin embargo, la estabilización de la proteína y el reclutamiento de coactivadores son las etapas reguladas por oxígeno.

La regulación transcripcional mediada por HIF requiere el reclutamiento de diversos coactivadores, que junto con HIF $\alpha$  y HIF $\beta$  forman un complejo transcripcionalmente activo. Estos coactivadores tienen como función estabilizar el complejo de iniciación de la transcripción y también acetilan histonas para permitir la apertura de la cromatina al paso de la maquinaria de transcripción. El principal complejo coactivador es CBP/p300 que interacciona con el dominio CTAD de HIF $\alpha$  ayudando al posterior reclutamiento del resto de coactivadores (Freedman, Sun et al. 2003).

HIF1 $\alpha$  es capaz de reclutar a las deacetilasas SRC1 y TIF2, miembros de la familia p160, que aumentan la capacidad transcripcional del complejo HIF en hipoxia (Carrero, Okamoto et al. 2000). Así mismo, Ref1 fomenta la transcripción mediada por HIF, favoreciendo la interacción de la subunidad alfa con cofactores como CBP/p300 (Carrero, Okamoto et al. 2000), mediante reacciones redox (Huang, Arany et al. 1996).

HIF $\alpha$ , como ya hemos mencionado antes, presenta dos dominios de transactivación, C-TAD y N-TAD y la principal diferencia funcional viene dada por su regulación mediante mecanismos de hidroxilación. El dominio N-TAD contiene un residuo de Prolina susceptible de hidroxilación (pro 564) que pertenece al dominio de degradación dependiente de oxígeno (ODD), de modo que es prácticamente imposible discriminar entre regulación transcripcional de este dominio y la regulación dependiente de la estabilidad de la proteína. A pesar de que hay trabajos que argumentan una regulación por hidroxilación del NTAD independiente de estabilidad, esta regulación es muy minoritaria en comparación con la de CTAD (Jiang, Zheng et al. 1997). Recientemente se han caracterizado los diferentes papeles transcripcionales que juegan NTAD y CTAD (Dayan, Roux et al. 2006) y concluyen que es la región CTAD la que determina la transcripción de la mayoría de las dianas de HIF, siendo el dominio más potente de transactivación, sin embargo, existe un conjunto de genes específicos que se regulan vía NTAD, apuntando una vez más a diferencias en la especificidad de ambos dominios. Por otro lado, se ha visto que las diferencias de especificidad de sustrato existentes entre HIF1 $\alpha$  y HIF2 $\alpha$  pueden venir mediadas por la unión de NTAD a diferentes cofactores transcripcionales (Hu,



Satur et al. 2007). Entre ambos dominios existe un dominio de inhibición, ID, que reprime la actividad transcripcional de CTAD en condiciones de normoxia (O'Rourke, Tian et al. 1999).

Mientras que la estabilidad de HIF se regula por la hidroxilación de sus dos prolinas en el ODD, la actividad de CTAD se regula por la hidroxilación de un residuo conservado de asparragina por medio de la actuación de una hidroxilasa llamada FIH (Lando, Peet et al. 2002) que, al igual que las EGLNs, es una dioxigenasa dependiente de 2-oxoglutarato y hierro. La hidroxilación en este residuo previene totalmente la unión de CTAD a los coactivadores transcripcionales CBP/p300, inhibiendo la actividad transcripcional de HIF (Lando, Peet et al. 2002); de hecho, el cambio de esta asparragina por otro aminoácido permite la unión de CTAD con CBP/p300, indicando la importancia de la hidroxilación de este residuo para el control transcripcional en normoxia.

#### II.4.1. OTRAS MODIFICACIONES QUE REGULAN A HIF.

Además de la regulación dependiente de oxígeno llevada a cabo por FIH y las EGLNs, existen otra serie de mecanismos de control de la estabilidad y la actividad transcripcional de HIF que implican modificaciones postaduccionales. Así, se han descrito modificaciones que afectan a la estabilidad de HIF $\alpha$  como es su unión a la proteína kinasa C (RACK1), que compete con la chaperona Hsp90 encargada de estabilizarla, promoviendo la ubiquitinación de HIF $\alpha$  y su degradación vía proteosoma (Liu, Baek et al. 2007). También se ha descrito recientemente que la interacción de HIF $\alpha$  con las proteínas Hsp70 y la proteína de interacción con Hsc-70 promueven su ubiquitinación y degradación (Luo, Zhong et al. 2010). Así mismo, se han descrito otras modificaciones como nitrosilaciones (Sumbayev, Budde et al. 2003), SUMOilaciones (Berta, Mazure et al. 2007) y acetilaciones (Jeong, Bae et al. 2002), cuyo efecto sobre la transcripción y estabilidad de HIF aún no se ha clarificado.

Modificaciones en el dominio CTAD de HIF $\alpha$  regulan la capacidad transcripcional del complejo; la fosforilación en un residuo de treonina de HIF $\alpha$  llevada a cabo por la casein kinasall permite aumentar la interacción de CTAD con p300 ya que impide la unión de FIH (Sang, Stiehl et al. 2003) y la activación de MAPK incrementa de actividad de CTAD (Sang, Stiehl et al. 2003). Adicionalmente se ha visto que fosforilaciones en la región ID reducen la actividad transcripcional de HIF impidiendo la translocación de HIF $\alpha$  al núcleo (Mylonis, Chachami et al. 2006) y que la fosforilación en el dominio de degradación dependiente de oxígeno (ODD) de

HIF $\alpha$  llevada a cabo por GSK3, contribuye al mantenimiento de la estabilidad de la proteína durante hipoxia crónica (Sodhi, Montaner et al. 2001).

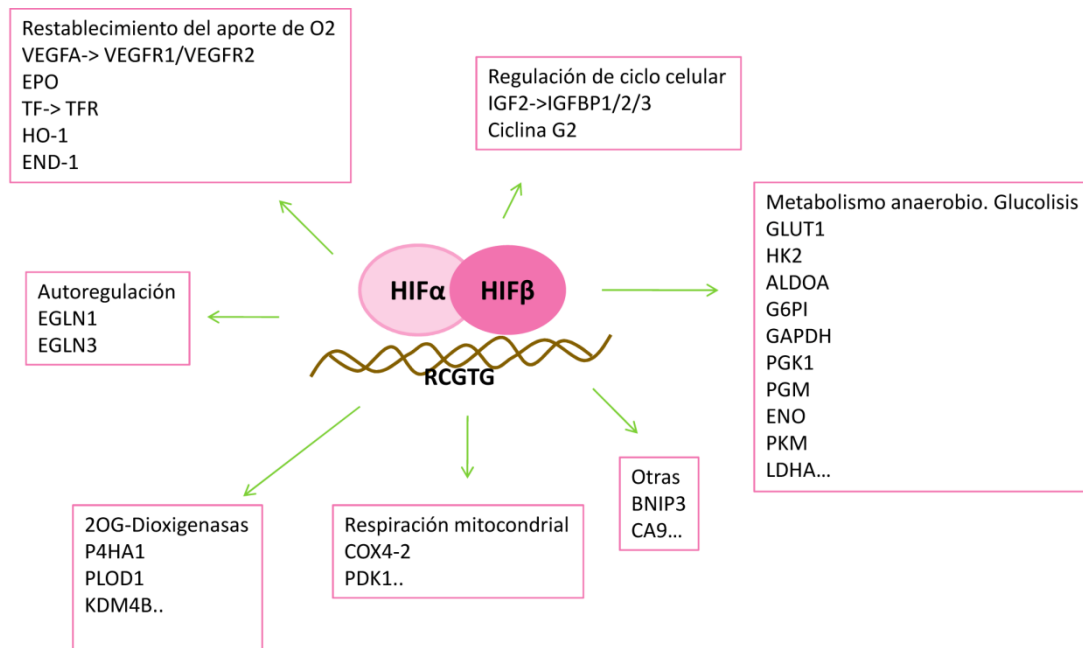
Por otro lado, se han descrito diferentes mecanismos que incrementan la transcripción de HIF $\alpha$  en determinados tipos celulares. Estímulos como la insulina (Zelzer, Levy et al. 1998), IGF 1/2 (Feldser, Agani et al. 1999), EGF (Zhong, Chiles et al. 2000) y Angiotensina II (Richard, Berra et al. 2000) entre otros, aumentan los niveles de proteína de HIF $\alpha$ , que es capaz de formar un complejo transcripcionalmente activo en condiciones de normoxia.

Así pues, todas estas modificaciones parecen jugar un papel en la regulación de HIF, sin embargo, se desconoce su importancia relativa en comparación con la regulación dependiente de hidroxilación.

## **II.5. TRANSCRIPTOMA DE HIF**

Las células responden a cambios en la disponibilidad de oxígeno activando una respuesta transcripcional coordinada mediada por HIF, que permite la adaptación a hipoxia. Esta respuesta conlleva la expresión de múltiples genes implicados en diversas funciones como son (fig 6) (Jiang, Zheng et al. 1997):

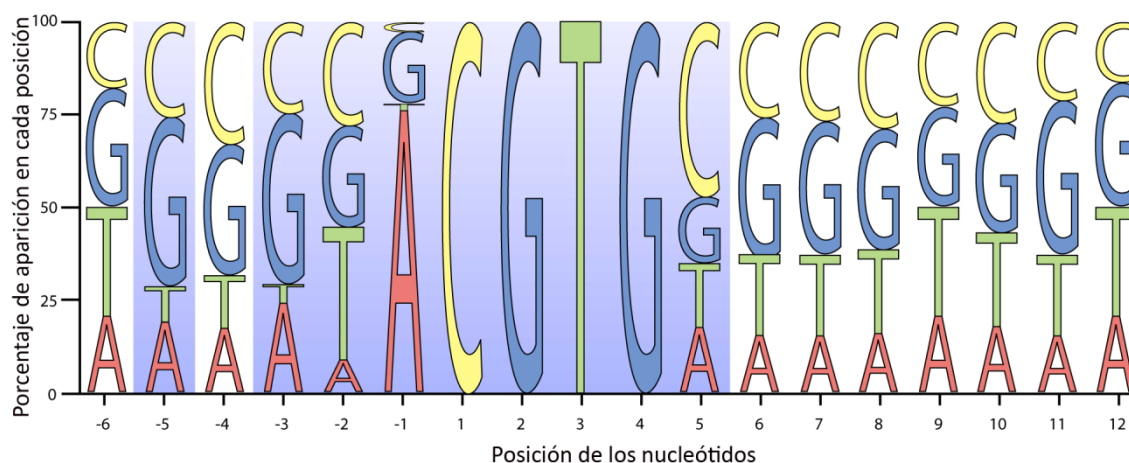
- Adaptación metabólica a la bajada de oxígeno que permite reducir el consumo de oxígeno y mantener los niveles de ATP constantes. Para ello, se inducen genes implicados en metabolismo anaerobio como GLUT1, HK2, GAPDH, etc, y respiración mitocondrial como COX4-2 y PDK1.
- Angiogénesis (VEGF), vasodilatación (iNOS), producción de eritrocitos (EPO) y transporte de hierro (transferrina) , que permiten restablecer el porte de oxígeno a los tejidos.
- Regulación de la propia ruta de HIF mediante la inducción de las EGLNs, EGLN1 y EGLN3.
- Remodelación de cromatina induciendo KDM4B, KDM3A.
- Regulación de la proliferación y diferenciación celular, ciclina G2, IGF2, entre otras funciones.



**Fig 6: Genes regulados por HIF.** Se han descrito más de 100 genes diana de HIF implicados en diversas funciones biológicas que permiten la adaptación celular a la bajada de oxígeno.

Actualmente se conocen más de 100 genes regulados por HIF (Wenger, Stiehl et al. 2005), pero aún estamos lejos de conocer la respuesta global a hipoxia. La aparición de nuevas técnicas de alto rendimiento como los perfiles de expresión génica (microarrays) han aportado gran cantidad de información acerca de los genes modulados por hipoxia, sin embargo, este tipo de aproximaciones no permite discernir entre dianas de HIF o de otros factores que se modulen por hipoxia, ni entre dianas directas o indirectas de HIF. Con el fin de subsanar estas carencias experimentales se han desarrollado técnicas moleculares específicas como son los ensayos de Chip-Chip, con los que se pretende identificar los elementos de respuesta a hipoxia, HREs, a los que se une HIF en la totalidad del genoma.

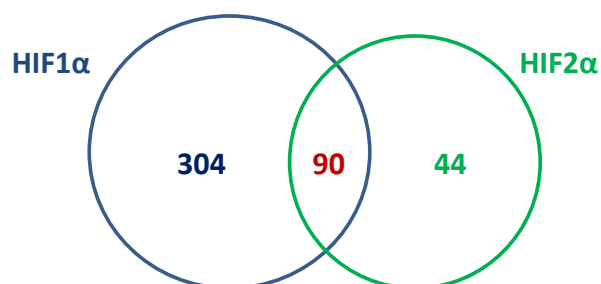
Trabajos enfocados al estudio individual de genes diana de HIF permitieron establecer la secuencia consenso de unión de HIF HRE, RCGTG. En nuestro proyecto, así como en otros trabajos anteriores (Pescador, Cuevas et al. 2005) (Wenger, Stiehl et al. 2005), se pone de manifiesto la importancia de las posiciones adyacentes al núcleo central del HRE a la hora de hacer distinciones entre HREs funcionales y no funcionales. En la figura 7 queda reflejada la frecuencia de aparición de cada base en cada posición en torno al RCGTG, así vemos que existen bases que aparecen con mayor frecuencia que la esperada por azar. Considerando por tanto que la secuencia HRE es la comprendida en las 18 pares de bases que contienen al RCGTG.



**Fig 7: secuencia consenso del HRE.** En el gráfico quedan reflejadas las posiciones en torno al núcleo central del HRE, RCGTG, y la frecuencia de aparición de cada base nucleotídica en cada posición (Wenger, Stiehl et al. 2005).

Por otro lado, se ha visto que la presencia de este núcleo central o HRE es necesaria pero no suficiente para la transcripción de los genes diana de HIF en respuesta a hipoxia. Existen una serie de regiones en torno al HRE necesarias para su actividad transcripcional a las que se unen otros factores de transcripción que colaboran con HIF, y que no necesariamente se inducen por hipoxia (Ebert and Bunn 1998). Así, estos factores pueden, desde estar potenciando la actividad transcripcional de HIF, hasta determinando regulaciones específicas de tejido.

Estudios de Chip-Chip para las dos isoformas de HIF $\alpha$  revelaron que ambas se unen a la misma secuencia central del HRE, sin embargo el conjunto de genes a los que se unen solapan en un bajo porcentaje (fig 8), siendo la subunidad 1 $\alpha$  la que modula la transcripción de un mayor número de genes en respuesta a hipoxia (Mole, Blancher et al. 2009). Así, a pesar de que HIF1 $\alpha$  y HIF2  $\alpha$  presentan algunos genes diana comunes como VEGF, ADM, GLUT1, NDRG1, etc, el conjunto de genes que modulan cada una tiene poco solapamiento con el de la otra. La subunidad HIF1 $\alpha$  es la encargada de regular la transcripción de todas las enzimas glucolíticas (Hu, Wang et al. 2003) y genes como BNIP3, PGK1 y las proteínas proangiogénicas HO1 y IL-8 entre otros. HIF2 $\alpha$  por su parte transcribe genes como EPO en tejido hepático, Oct-4, implicado en autorenovación en células madre, Ciclina D, el factor TGF $\alpha$  y DMT1, transportador iónico intestinal, entre otros (Covello, Kehler et al. 2006).



**Fig 8: Número de secuencias RCGTG a las que se unen HIF1α y HIF2α en respuesta a hipoxia.** Resultado de las regiones de unión de ambas subunidades en ensayos de Chip-Chip en células MCF7. Adaptado de (Mole, Blancher et al. 2009).

### II.5.1. COFACTORES TRANSCRIPCIONALES

HIF1α y HIF2α son muy similares en cuanto a la estructura de sus dominios, su regulación por hipoxia, su mecanismo de unión DNA y a la activación de la transcripción, sin embargo, tal y como muestran estudios bioquímicos (Mole, Blancher et al. 2009) y modelos knockout in vivo, sus funciones no son redundantes. Actualmente no se conocen a fondo los motivos subyacentes a esta diferencia funcional, sin embargo la respuesta más plausible es la especificidad de colaboración con determinados factores de transcripción por parte de cada isoforma (Hu, Sataur et al. 2007).

Así, existen una serie de cofactores que colaboran específicamente en la transcripción de genes diana con HIF1α, como AP1 (Damert, Ikeda et al. 1997) que colabora en la transcripción de VEGF, y CREB1 y ATF1 en la de LDHA (Firth, Ebert et al. 1995); o con HIF2α, como el modulador de NFκB NEMO, que promueve la transcripción mediada por HIF2α en normoxia (Bracken, Whitelaw et al. 2005). También se ha visto que el factor Ets-1 interacciona con HIF2α en la transcripción del receptor 2 de VEGF (Elvert, Kappel et al. 2003) y, otro miembro de esta misma familia llamado Elk1, coopera con HIF2α para transcribir genes como EPO, CITED2, IGF3 y PAI1 (Aprelikova, Wood et al. 2006) (Hu, Sataur et al. 2007).

También existen cofactores tejido específicos que son necesarios para la transcripción de algunos genes, es el caso de los cofactores HNF4 y Smad3. HNF4 es un cofactor específico de los tejidos hepático y renal, que es requerido para la transcripción del gen que codifica para la eritropoyetina EPO, para ello se une a dos sitios próximos a su HRE (Zhang, Tsuchiya et al. 1999). Smad3 interacciona con HIF1α y se une a un sitio próximo al HRE del gen de VEGF

umentando su tasa de transcripción (Sanchez-Elsner, Botella et al. 2001), se ha visto que Smad3 interacciona con CBP/p300 de modo que el mecanismo por el cual actúa sería estabilizando el complejo HIF1 $\alpha$ -CBP/p300. Así mismo se ha visto que para la transcripción de VEGF es importante la unión del cofactor STAT3 en cánceres de próstata y páncreas (Gray, Zhang et al. 2005).

#### II.5.2. INTERACCIÓN DE HIF $\alpha$ CON OTRAS RUTAS CELULARES

Por último mencionar que HIF $\alpha$ , una vez que se estabiliza por hipoxia puede estar participando en otras vías de señalización que no implican la unión a su secuencia consenso HRE, como son las rutas de Myc y Notch. La capacidad proliferativa de las células durante procesos hipóxicos viene dado en parte por la interrelación entre HIF y el pro-oncogén c-Myc, así se ha descrito que la participación de HIF1 $\alpha$  o HIF2 $\alpha$  en la transcripción de genes diana de c-Myc tiene efectos opuestos, 1 $\alpha$  inhibe la modulación de sus genes diana mientras que 2 $\alpha$  la potencia (Gordan, Thompson et al. 2007).

Por otro lado, se sabe que la hipoxia promueve el estado indiferenciado en células madre y esta función requiere la activación de la cascada de señalización de Notch, así se ha descrito que Notch es capaz de reclutar a HIF para unirse a los promotores de los genes diana de Notch mediando su transcripción en condiciones de hipoxia (Gustafsson, Zheng et al. 2005).

#### II.6. IDENTIFICACIÓN DE DIANAS DE HIF.

A lo largo de mucho tiempo se han ido identificando numerosas dianas de HIF como resultado de trabajos realizados en genes individuales. Sin embargo el desarrollo de nuevas aproximaciones experimentales, como las técnicas de alto rendimiento, ha acelerado la identificación a gran escala de genes regulados por HIF. Además de las aproximaciones experimentales, existen técnicas computacionales que nos permiten predecir sitios de unión de factores de transcripción. Estas técnicas junto con la secuenciación de genomas completos ofrecen una alternativa o complemento a las técnicas experimentales. En relación al factor inducible por hipoxia HIF, hasta el inicio de este proyecto nadie había aplicado este tipo de abordajes computacionales con el fin de conocer la totalidad de sus genes diana. El conocimiento global de la regulación mediada por HIF en hipoxia puede permitir el desarrollo

de nuevas técnicas farmacológicas aplicadas a enfermedades que cursan con hipoxia, de ahí la importancia de ampliar el conocimiento existente sobre el transcriptoma de este factor.

#### II.6.1. TÉCNICAS COMPUTACIONALES

En los últimos años se han desarrollado numerosas aproximaciones computacionales que permiten estudiar la regulación mediada por un determinado factor de transcripción. Estas estrategias requieren el análisis, procesamiento e integración de una gran cantidad de información génica y una posterior validación experimental.

La identificación computacional de elementos reguladores es un proceso tedioso, sin embargo existen una serie de criterios que lo facilitan. A partir del conocimiento de un conjunto de secuencias de unión de genes diana de un factor de transcripción puede desarrollarse un modelo que describa las características propias que debe tener una secuencia para que ese factor se una, lo que puede ser utilizado como modelo predictivo de nuevas secuencias (van Helden 2003). Una de las limitaciones de estos modelos es la escasa cantidad de elementos reguladores en *cis* que suelen estar validados para un factor de transcripción y otra es que, generalmente, solo tomamos estrictamente la secuencia de unión del factor de transcripción, lo que no es del todo correcto, pues las regiones adyacentes pueden determinar la funcionalidad de ese elemento regulador debido a la unión de otros factores (Davidson, Rast et al. 2002). Así pues, del alineamiento de estas secuencias conocidas podemos extraer un patrón común o consenso del elemento de unión de nuestro factor. Con estas secuencias consenso se pueden elaborar matrices de frecuencia (PWM) que asignan un valor numérico a cada base para cada posición de la secuencia en función de su frecuencia de aparición en cada posición (Roulet, Fisch et al. 1998). Así la comparación de nuevas secuencias localizadas con esta PWM nos permite asignar una puntuación a la misma, lo que es indicativo de su parecido con secuencias funcionales.

A la hora de seleccionar las secuencias reguladoras a estudiar, existen una serie de filtros que permiten aumentar la eficiencia de obtención de elementos reguladores funcionales disminuyendo los falsos positivos (Blanchette and Sinha 2001) (Sinha, van Nimwegen et al. 2003). Uno de ellos es realizar la búsqueda de secuencias reguladoras solo en regiones no codificantes, ya que en ellas es donde se localizan prácticamente la totalidad de las regiones reguladoras (Little 2005; Elgar and Vavouri 2008; Visel, Rubin et al. 2009). Otro de los criterios a tener en cuenta a la hora de identificar nuevas secuencias reguladoras es el uso

## Introducción

de secuencias ortólogas, método denominado como “huella filogenética” (*Phylogenetic Footprinting*), esto introduce un filtro evolutivo suponiendo que, las regiones con alguna funcionalidad acumulan mutaciones con menor frecuencia que las regiones no funcionales y por tanto están conservadas evolutivamente. Estos criterios suponen un incremento de la especificidad cerca del 50% a la hora de predecir elementos reguladores reales (Tagle, Koop et al. 1988).

La aplicación de técnicas computacionales, por sí solas, para la identificación de elementos reguladores presenta algunas carencias, así, la utilización de PWMs como único criterio para predecir sitios de unión de un factor de transcripción no es muy efectivo, puesto que el genoma tiene un elevado tamaño y además presenta numerosas regiones reguladoras silenciadas. Así pues, la propuesta de combinar aproximaciones computacionales y experimentales en un único modelo predictivo, parece ser la propuesta más fiable.



## ***OBJETIVOS***



Conocer el transcriptoma de HIF es esencial para comprender de forma global la respuesta celular a hipoxia. Por ello nos planteamos los siguientes objetivos a desarrollar en nuestro proyecto:

1. Obtener un listado fidedigno de los potenciales genes diana de HIF de todo el genoma mediante la elaboración de una estrategia computacional que combina datos experimentales provenientes de perfiles de expresión génica y la identificación *in silico* de los sitios de unión de HIF (HBSs) en todos los genes del genoma.
2. Validar el listado de potenciales genes diana de HIF predichos por nuestra estrategia bioinformática mediante técnicas experimentales.
3. Estudiar las funciones biológicas más representadas en la respuesta celular a hipoxia, implicadas tanto en la adaptación celular a este estímulo como en procesos patológicos.
4. Identificar SNPs que se localicen en los sitios de unión de HIF (HBSs) de los potenciales genes diana identificados, su presencia podría alterar la respuesta transcripcional mediada por este factor.



## ***MATERIALES Y MÉTODOS***



#### IV.1. CULTIVOS CELULARES Y TRATAMIENTO DE HIPOXIA

Las líneas celulares Hela y c2c12 se crecieron en medio DMEM (Sigma) y las líneas Hepa C1/C4 en medio MEM  $\alpha$ . En todos los casos los medios de cultivo se suplementaron con 100U/ml de penicilina, 100 $\mu$ g/ml de estreptomina y 10% (v/v) de suero fetal bovino. Para el proceso de diferenciación de las células c2c12, éstas se cultivaron hasta un 80% de confluencia y se sustituyó el medio a DMEM suplementado con 2% de suero de caballo durante tres días. Las células se crecieron a 37°C en una atmósfera humidificada al 5% de CO<sub>2</sub>. Para los tratamientos de hipoxia, las células se expusieron a un 1% de oxígeno en la cabina Whitley de hipoxia (don Whitley Scientific, UK) durante los tiempos indicados.

#### IV.2. GENERACIÓN DE PLÁSMIDOS

La construcción reportera de MIF se generó a partir de DNA genómico extraído de células Hela, a partir del cual se amplificó la región promotora usando los primers 1+2 (Tabla2). El producto de amplificación de la PCR fue clonado en el vector pCR2.TOPO (Invitrogen) y posteriormente fue subclonado en el vector reportero pGL3Basic (Invitrogen) en los sitios BgIII y XhoI. Las mutaciones del HRE fueron realizadas usando el kit PCR Quick Change Site-Direct mutagenesis kit (Stratagene), con los primers 3+4 y 5+6. La identidad de todas las construcciones fue verificada por secuenciación.

Numero	Nombre	Secuencia	Posición
1	HREwt MIF F	CTCGAGCGGCGGTGGCGTCACAAAAG	-31
2	HREwt MIF R	AGATCTCGGCATGATGGCAGAAGGACC	108
3	HREmut MIF F	AAGTCAGGCTAGCAGCTCAGCGGC	NA
4	HREmut MIF R	GCCGCTGAGCTGCTAGCCTGACTTC	NA
5	HRE snp MIF F	AAGTCAGGAACGTAGCTCAGCGGC	NA
6	HRE snp MIF R	GCCGCTGAGCTACGTTCTGACTTC	NA

Tabla 2: Primers utilizados para el clonaje de los promotores de MIF y GYS1 y la mutagénesis de sus HREs.

### **IV.3. ENSAYOS REPORTEROS**

Los ensayos reporteros fueron realizados en células Hela que se sembraron 24 horas antes de la transfección en placas de seis pocillos. Las transfecciones se realizaron usando 2  $\mu$ l lipofectamina (Invitrogen) junto con una mezcla de 1  $\mu$ gr de DNA total por pocillo. Las mezclas de DNA de transfección contenían 0,9  $\mu$ gr del plásmido reportero indicado y 0,1  $\mu$ gr de un plásmido que codifica para el enzima luciferasa-Renila bajo el control del promotor SV40. Tras cuatro horas de transfección se puso medio fresco a las células y 12 horas después se replaquearon en placas de 24 pocillos. Una vez adheridas, las células fueron expuestas a condiciones de normoxia (21%O<sub>2</sub>) e hipoxia (1%O<sub>2</sub>) durante 12 horas. Finalmente las actividades renila y luciferasa fueron medidas por duplicado usando un sistema dual de medición (Promega, Madison, WI, USA). A modo de corregir la eficiencia de transfección, la actividad luciferasa fue normalizada frente a la actividad renila-luciferasa.

### **IV.4. INMUNOPRECIPITACIÓN DE CROMATINA (ChiP-Qpcr-RT)**

Para los ensayos de ChiP se utilizaron células Hela que fueron cultivadas en placas de 10cm hasta que alcanzaron una confluencia del 80%, en ese punto se mantuvieron en condiciones de normoxia (21% O<sub>2</sub>) o hipoxia (1% O<sub>2</sub>) durante 6 horas. Tras ese tiempo fueron fijadas durante 12 minutos a 4°C añadiendo en el medio de cultivo formaldehído a una concentración final del 1% (v/v), la fijación se paró añadiendo glicina a una concentración final de 0,125M. Posteriormente las células se lavaron con PBS a 4°C y fueron lisadas en 1ml de tampón de lisis (1% SDS, 10mM EDTA, 50mM Tris/HCl, pH 8, 1 y un cocktail de inhibidores de proteasas, Roche) durante 10 minutos a 4°C. A continuación los lisados se sonicaron para fragmentar el DNA en tamaños entre 200 y 1500 pares de bases, se comprobó el tamaño de fragmentos del sonicado corriendo las muestras de normoxia e hipoxia en un gel de agarosa al 2%. El paso de sonicación es crítico para una inmunoprecipitación eficiente. Tras eliminar los restos insolubles por centrifugación, 50  $\mu$ l de cada lisado se apartaron y fueron guardados a -20°C (input), mientras que 100  $\mu$ l de muestra se diluyeron en 1ml de solución de inmunoprecipitación (1% Triton X-100, 2mM EDTA, 150mM NaCl y 20mM Tris/HCl, pH 8.1). Los lisados fueron tratados con 200  $\mu$ gr de DNA de esperma de salmón /proteína A agarosa (Upstate Biotechnology, Lake Placid, NY, USA) durante 1 hora a 4°C para eliminar uniones inespecíficas. Posteriormente las muestras fueron inmunoprecipitadas dos veces a 4°C, inicialmente con suero total de conejo durante 6 horas (Inmunoglobulina G, IgG control) y luego 12 horas con un anticuerpo monoclonal anti-HIF1 $\alpha$  (Abcam, ab2185). Los



inmunocomplejos se precipitaron usando una mezcla de 400µgr de DNA de esperma de salmón/Proteína A agarosa. Seguidamente se lavaron durante 15 minutos secuencialmente con las siguientes soluciones: TSE I (0.1% SDS, 1% Triton X-100, 2mMEDTA, 20mM Tris/HCl, pH 8.1 y 150mM NaCl), TSEII (0.1% SDS, 1% Triton X-100, 2mM EDTA, 20mMTris/HCl, pH 8.1 y 500mM NaCl) y buffer III (0.25M LiCl, 1% NP-40, 1% deoxicolato, 1mMEDTA y 10mM Tris/HCl, pH 8.1). Finalmente los complejos se lavaron dos veces con TE (10mM Tris, pH 8.0 y 1mM EDTA) y se eluyeron, separando las Proteínas A de los inmunocomplejos, usando un buffer que contenía 1% SDS y 0.1M NaHCO<sub>3</sub>. Desechamos la Proteína A-agarosa y nos quedamos con el sobrenadante que incubamos durante una noche a 65°C con un 200mM NaCl (concentración final) para romper las uniones DNA-proteína. Las proteínas se eliminaron añadiendo proteinasa K (30 mg/muestra) durante 2 horas a 42°C y el DNA fue extraído con fenol-cloroformo y precipitado con etanol. El DNA inmunoprecipitado se amplificó por qPCR usando los primers (7-42) indicados en la tabla3. El sistema empleado fue *StepOnePlus™ Real-Time PCR System* (Applied Biosystems)

#### IV.5. EXTRACCIÓN DE RNA Y q-PCR-RT.

Las células se lisaron en 1ml de Ultraspect (Biotecx) y se extrajo el RNA total por fenol-cloroformo y precipitación con etanol al 70%. El RNA se analizó y que fue cuantificado, posteriormente se retrotranscribió 1µgr a cDNA (Improm-II transcriptasa reversa, Promega) usando “random primers”. 1µl de cDNA fue usado como molde para la amplificación llevada a cabo mediante el sistema de qPCR Light Cyclor (Roche Applied Science) usando los primers que se indican en la tabla 4. Los datos fueron analizados con el software Light Cyclor versión 3.5.28 (Idaho Technology). Cada muestra fue analizada por duplicado y se calculó la cantidad relativa de cada mRNA mediante la técnica de  $\Delta\Delta CT$  usando  $\beta$ -actina como gen de referencia.

	Nombre	Secuencia	Posición
7	UGP2 Chip HRE1 F	GGGAGGGGGCAGCTAGGG	786
8	UGP2 Chip HRE1 R	GCGGACCTACAGCGGCTATTT	886
9	UGP2 Chip HRE2 F	CTTAGTGTCTCCGCTTCTA	1664
10	UGP2 Chip HRE2 R	GCTACTCCAGTTTCGCAATA	1744
11	GBE1 Chip F	ACAAACGCTCACACATACACA	44901
12	GBE1 Chip R	TCAGCAGAGCATGTAGCAAA	44708
13	CARHSP1 Chip F	CAGCCGCTTCGCTCCGAC	18
14	CARHSP1 Chip R	CCGCTCTCCCGCCAGTCTAAG	178
15	IDH2 Chip F	TGACGTGGGCCTGGTATCA	712
16	IDH2 Chip R	CCCCTGGAACGAGACGAGCAC	545
17	SERTAD2 Chip F	CGCCAGCAATCCGAGCCTTA	693
18	SERTAD2 Chip R	GCCTGGAGCTGTCGCCTCTAC	593
19	MKRN1 Chip F	G TTCACACGCCACGCAAGGTT	25074
20	MKRN1 Chip R	GCTGGGCGAGATGTTGCTTAT	24932
21	RPLP0 Chip F	TGAAGAGCAGAGGCGACCCAC	513
22	RPLP0 Chip R	ATGGGTGTCGGCGTGAC	335
23	MAFF Chip F	GGCGCTGTGCTCTACGTGAC	-514
24	MAFF Chip R	GCACCGGGCAGGATACC	-647
25	TPD52 Chip F	CCTCGGCCTGGAGCGCTTTC	-3
26	TPD52 Chip R	AGTCCATCTCCTCGCCGCCT	-176
27	MIF F	GTCACAAAAGGCGGGACCACA	-15
28	MIF R	CAGGAGACCCGCGCAGAG	82
29	CDC2L6 F	GCACAAACCTCGCAGTG	205052
30	CDC2L6 R	ATCCTTTCTAAACTGCCTAGT	204892
31	Chip EGLN3+_F	CGTGGAGGACTGGCTCTAAG	12638
32	Chip EGLN3+_R	GGTGTGCTCGGGTGTG	12543
33	Chip EGLN3-_F	CCCTTAACGTTGACTTTCGCTC	876
34	Chip EGLN3-_R	ACGGGAGGCACTCGGAG	725
35	NEG1 F (IRS4)	AGCGGGCAAACAACACGTGA	-212
36	NEG1 R (IRS4)	CTACTGGTGGCCGCTCG	-119
37	NEG3 F (STT3A)	GCAAGGGCCTATTTACAGCGTA	372
38	NEG3 R (STT3A)	GAGCGCGAAAGAACGTG	219
39	NEG7 F (HIVEP1)	CGGTTGCCATGGAGAGGTAGT	48463
40	NEG7 R (HIVEP1)	GTGCACAGTCACGTTACCAGC	48636
41	NEG8 F (LTBP1)	TCGCACGCCTTCCCAATTAG	-498
42	NEG8 R (LTBP1)	GGTGGCCGAGAGCGAG	-336

**Tabla 3: Primers utilizados para la amplificación del DNA inmunoprecipitado por ChiP.** En la tabla se muestra el número asignado, la nomenclatura dada al primer, la secuencia y la posición de hibridación del mismo con respecto al inicio del gen.

Número	Nombre	Secuencia	Posición
43	GYS1 F	TTCTACAACAACCTGGAG	11097
44	GYS1 R	CTGAGCAGATAGTTGAGC	13483
45	UGP2 F	CAGAGACCTCCAGAAGATTCG	505
46	UGP2 R	GTTCAACACAGAAGATATGTTATCAGG	588
47	GBE1 F	CTTCCGTTTTGATGGTGTCA	1125
48	GBE1 R	AGTCACCCGAAAAACCTTGA	1195
49	PPP1R3C F	CTGAGTGCGAAGTTGCTCAG	31
50	PPP1R3C R	TGGATCTAGCACATGGATATT	123
51	Bactina F	CCCAGAGCAAGAGAGG	2055
52	Bactina R	GTCCAGACGCAGGATG	1250

**Tabla 4: Primers para la amplificación de los cDNAs de los genes indicados.** En la tabla se muestra el número asignado, la nomenclatura dada al primer, la secuencia y la posición de hibridación del mismo con respecto al inicio del gen.

#### IV.6. ANÁLISIS DE LOS DATOS DE LOS PERFILES DE EXPRESIÓN DE LA BASE DE DATOS GEO

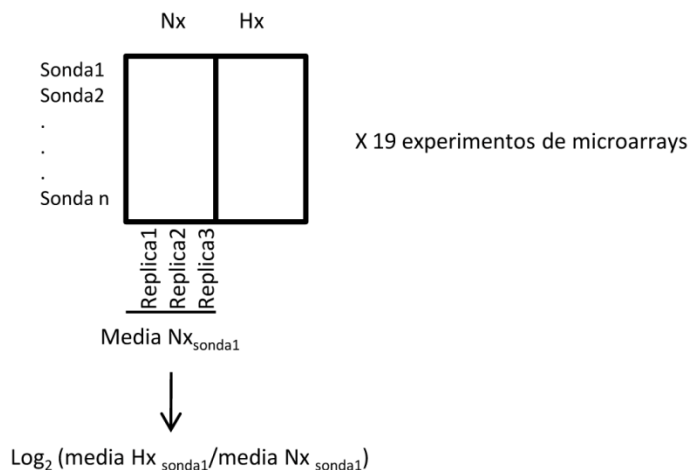
De la base de datos Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) se extrajeron 16 experimentos de perfiles de expresión génica (microarrays) de células expuestas a condiciones de hipoxia, miméticos de la hipoxia (DMOG y DFO) o sobreexpresando las subunidades HIF1 $\alpha$  o HIF2 $\alpha$ , todos los experimentos contenían además el perfil de expresión génica de células en normoxia como referencia. Alguno de estos experimentos englobaba más de un tratamiento, como por ejemplo, hipoxia y sobreexpresión de HIF, para estos casos se generaron dos grupos de datos. De este modo, generamos 19 tablas de datos a partir de 16 experimentos de GEO independientes comparando cada tratamiento con el control de células en normoxia (tabla 5).

GEO_ID	Tipo celular	Tratamiento
GDS1648	Hepatocitos	Hipoxia 8h
GDS1772	RCC4	Hipoxia 18h
GDS1807	Línea humana de linfoma de Burkitt's, P493-6	Hipoxia 0.1%O <sub>2</sub> 29h
GDS2036	Macrofagos derivados de monocitos de 4pacientes sanos	Hypoxia 24 h
GDS2018	HEK293T	HIF1a 1%O <sub>2</sub> 16h
GDS2018	HEK293T	HIF2a 1%O <sub>2</sub> 16h
GSE2020	HEK293T	Hipoxia 1%O <sub>2</sub> 16h
GSE3045	Astroцитos	Hipoxia 1%O <sub>2</sub> 24h
GSE3051	HeLa	Hipoxia 1%O <sub>2</sub> 24h
GSE3188	MCF7	DMOG
GSE3188	MCF7	Hipoxia
GSE3195	MEF	Hipoxia 0.1%O <sub>2</sub> 6h
GSE3196	MEFs wt	Hipoxia 0.1%O <sub>2</sub> 6h
GSE4086	Línea humana de linfoma de Burkitt's, P493-6	Hipoxia 0.1% O <sub>2</sub> 29h
GSE4630	Macrofagos derivados de monocitos	0%O <sub>2</sub> 24h
GSE4725	Células humanas de musculatura lisa de aorta (HASMC)	Hipoxia 1%O <sub>2</sub> 16h
GSE5579	RNA de células endoteliales linfoides de 6 individuos	Hipoxia 1%O <sub>2</sub> 24h
GSE4725	U251	DFO 16h
GSE9234	HT29	Hipoxia

**Tabla5: Experimentos de expresión génica seleccionados a partir de la base de datos GEO Data Sets.** Seleccionamos 19 conjuntos de datos a partir de 16 experimentos independientes de microarrays. En la tabla aparecen distribuidas en columnas la información relativa al número de identificación del experimento, el tipo celular en el que ha sido realizado y el tipo de tratamiento y duración del mismo.

Con el fin de encontrar un conjunto de genes que se modulase de forma significativa en todo este grupo de experimentos, llevamos a cabo dos métodos de análisis de los mismos:

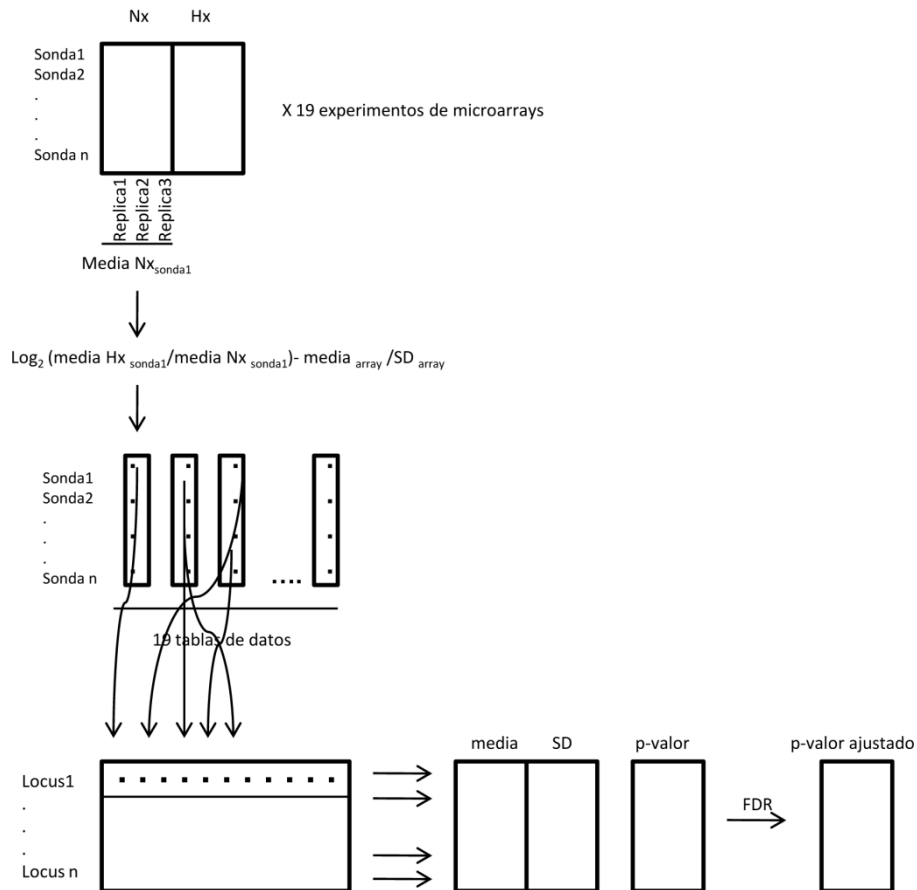
- Primero, para cada una de las tablas de datos se calculó la media de fluorescencia de las réplicas biológicas de cada sonda hibridada en la plataforma del microarray, descartando aquellas sondas con valores de fluorescencia no significativa respecto al background (hibridaciones “null”). Luego, se calculó el efecto del estímulo inferido (hipoxia, miméticos de hipoxia o sobreexpresión de la subunidad  $\alpha$ ), para ello se calculó el logaritmo del ratio de las medias de fluorescencia de las sondas de las muestras tratadas y sin tratar (condiciones de normoxia) (fig 9).



**Fig 9: Esquema del proceso de tratamiento de los datos de GEO para los análisis individuales.** De cada experimento de microarray calculamos los log ratio de las medias de fluorescencia en hipoxia frente a normoxia para cada sonda hibridada.

-Así mismo, realizamos un meta-análisis de los datos en el que cada locus fue tratado de forma independiente y se comparó con la hipótesis nula de que ningún gen se modulaba por el tratamiento ( $\log (\text{tratamiento}/\text{Nx})$  igual a cero). Primeramente estandarizamos los valores de los log-ratio de cada sonda en cada experimento, restándole a cada uno de ellos la media de los log-ratio de todas las sondas de dicho experimento y dividiéndolo por su desviación típica. Recopilamos el log-ratio normalizado de todas las sondas que hibridaban en un mismo locus, presentes en todos los experimentos de microarrays y, asumiendo una distribución normal de los datos, se comparó la media con la hipótesis nula de que no se indujesen por hipoxia (la media de los log-ratio de un gen que no se induce por hipoxia es cero) en un t-test simple. Finalmente los p-valores se corrigieron con un test de comparaciones múltiples usando *false discovery rate* (FDR) como método de corrección. Así consideramos que aquellos genes que presentaban unos p-valores ajustados por debajo de 0,01 estaban modulados por los distintos tratamientos (fig 10)

Para todo este análisis, se realizaron programas (“scripts”) en el lenguaje de programación Perl complementado con extensiones de lenguaje R para los cálculos estadísticos (anexo I).



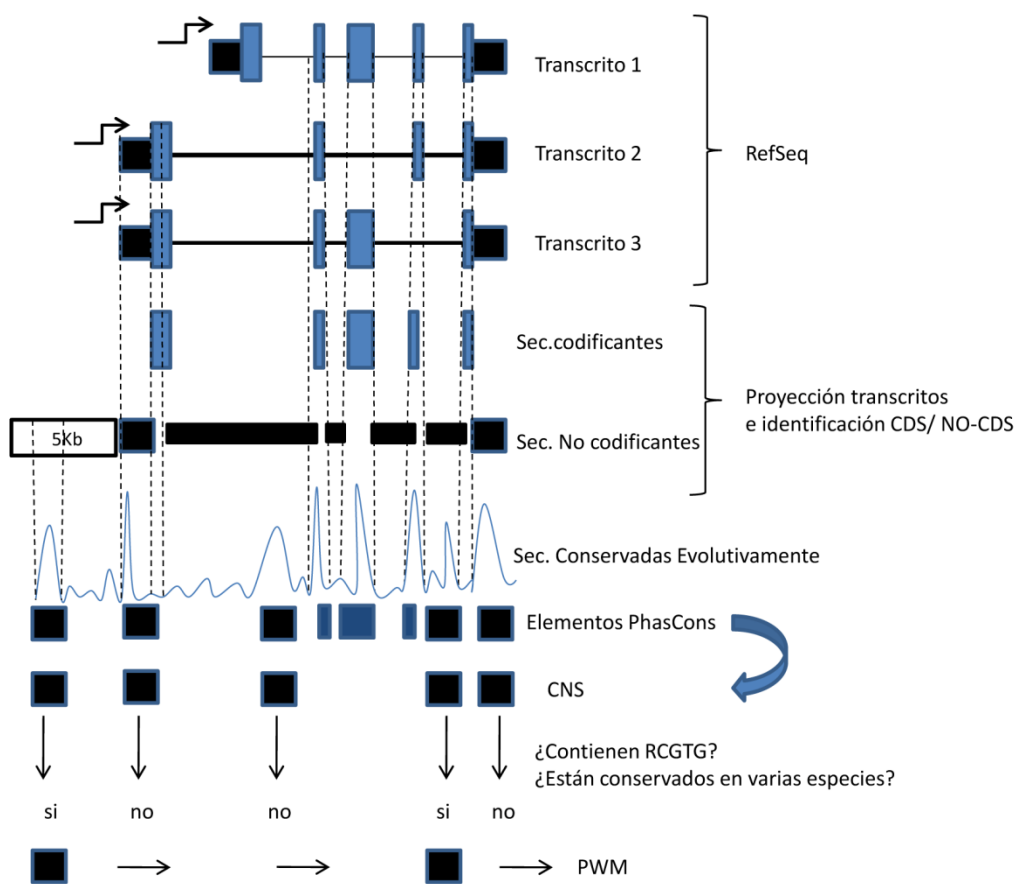
**Fig 10: Esquema del proceso de tratamiento de los datos de GEO para el meta-análisis.** De cada experimento de microarray extraímos los log-ratio para cada sonda y posteriormente se trataron en conjunto todas las sondas de las 19 tablas de datos referidas a un mismo locus, asignándoles un p-valor que finalmente ajustamos por un test de comparaciones múltiples (FDR).

#### IV.7. IDENTIFICACIÓN Y PUNTUACIÓN DE LOS SITIOS DE UNIÓN DE HIF (HBS)

Para la identificación de los sitios de unión de HIF restringimos nuestra búsqueda a las regiones no codificantes de los genes. Para ello, de cada locus génico tomamos todos los transcritos de referencia (“revisados”, RefSeq) del servidor del UCSC, identificamos las regiones no codificantes de los mismos y proyectamos todos los transcritos en una única secuencia, de modo que obtuvimos para un locus dado todas las posibles regiones no codificantes (fig11). Entre las regiones no codificantes se incluye un fragmento de 5Kb por encima del sitio inicial de la transcripción para cada gen. En el caso de existir solapamiento de esta región con otro locus vecino esta región se acortó para evitarlo, y en aquellos genes que

presentaban diferentes sitios de inicio de la transcripción, tomamos en todos los casos el más externo de ellos.

Una vez que extrajimos todas las regiones no codificantes procedimos a identificar dentro de ellas regiones conservadas evolutivamente. Para ello buscamos elementos PhasCons de vertebrados y mamíferos (Siepel, Bejerano et al. 2005) localizados dentro de regiones no codificantes. Aquellos elementos PhasCons muy próximos entre si fueron fusionados en uno solo, siempre y cuando al menos el 50% del resultado de la fusión estuviera conservado. Nos referimos a estos PhasCons dentro de regiones no codificantes como CNSs (secuencias no codificantes conservadas) (fig11).



**Fig 11: Esquema de la obtención de CNS para la identificación de sitios RCGTG.** Partimos de las secuencias RefSeq de los transcritos de todos los genes del genoma, las proyectamos en una única secuencia, y obtuvimos todas las posibles regiones no codificantes. En ellas localizamos elementos conservados evolutivamente (PhasCons) y extrajimos regiones no codificantes conservadas evolutivamente o CNSs. En ellas buscamos la secuencia del núcleo central del HRE (RCGTG) y comparamos estas secuencias con una matriz de frecuencias (PWM) para asignarles una puntuación.

Como resultado de este proceso obtuvimos un listado de coordenadas genómicas correspondientes a estas regiones conservadas evolutivamente localizadas en regiones no codificantes. Posteriormente, usando la herramienta “tables” del UCSC (<http://genome.ucsc.edu/cgi-bin/hgTables>) descargamos los alineamientos múltiples correspondientes a estas coordenadas. Finalmente, identificamos en los alineamientos correspondientes a estas CNSs sitios conservados de unión de HIF (RCGTG) y descartamos aquellas regiones cromosómicas que carecen de este motivo. Así mismo consideramos como potenciales sitios de unión de HIF solo aquellos en los que el motivo RCGTG estaba conservado en al menos cuatro especies, entre las que deben estar humano, ratón o rata, y alguno de los primates (fig 11).

Una vez obtenidos todos los sitios RCGTG conservados, se asignó un valor numérico que determinaba lo similares que eran nuestras secuencias con motivos RCGTG conservados a secuencias HREs funcionales. La asignación de este valor se dio por comparación con una matriz de frecuencias (PWM, *Position weight matrix*) generada a partir de 23 secuencias genómicas que contenían HREs bien caracterizados correspondientes a 22 genes diana de HIF conocidos, incluyendo para cada uno su ortólogo (anexo II). La comparación de la distribución de bases en estas posiciones frente a la esperada por azar mediante un test Chi-cuadrado, indicó que existían bases fuera del núcleo (RGTGC) del HRE que aparecían en esa posición con mayor probabilidad ( $p < 0,01$ ) que la esperada por azar; y esto ocurría entre las posiciones -1 y 17 con respecto al RGTGC (anexo II). Así, para cada una de estas posiciones se calculó el log-ratio de la frecuencia observada para cada nucleótido con respecto a la frecuencia esperada por azar (“background”). El background se calculó como la frecuencia de aparición de cada nucleótido en una posición dada a lo largo de todas las CNSs. Las frecuencias background obtenidas para cada base fueron: A= 0.275, C=0.233, G= 0.229 y T= 0.273. Así, estos log-ratios fueron posicionados en una matriz de 4 nucleótidos (columnas) X 18 posiciones (filas), siendo la puntuación para cada nucleótido  $i$  en una posición  $j$ :

$$S_{ij} = \log \left( \frac{frec_{ij}^{observados}}{frec_{i}^{background}} \right)$$



No todas las posiciones tenían el mismo peso o importancia a la hora de aportar información a la puntuación final de un HRE, por ello se tubo en cuenta el grado de contribución de cada posición. Para ello se calculó el contenido de información de cada posición  $j$  aplicando la ley de entropía de Shannon (Schneider, Stormo et al. 1986):

$$I_j = - \sum \text{frec } i^{\text{background}} * \log_2(\text{frec } i^{\text{background}}) - \sum \text{frec } ij^{\text{observados}} * \log_2(\text{frec } ij^{\text{observados}})$$

Siendo la puntuación final para cada secuencia el sumatorio de las puntuaciones individuales “equilibradas” de todos sus nucleótidos:

$$S = \sum_j I_j * S_{ij}$$

Toda la información genómica necesaria para esta aproximación la obtuvimos del servidor del UCSC (<http://genome.ucsc.edu/>). Las coordenadas corresponden a la versión hg18 del genoma humano. Los análisis fueron realizados por programas escritos en Perl (“scripts”) (anexo I).

#### IV.8. CLASIFICACIÓN DE LOS GENES IDENTIFICADOS EN: DIANAS/ NO DIANAS DE HIF.

Una vez obtenido el listado de potenciales genes diana de HIF, generamos un modelo probabilístico empírico que nos permitió clasificar a los genes en dianas o no de HIF. Este modelo tenía en cuenta varios parámetros a la hora de clasificar un gen, por un lado la magnitud de su modulación por hipoxia y el p-valor ajustado asignado a esta magnitud, que indicaba la significancia de esta modulación, ambos derivados del meta-análisis de los perfiles de expresión. Por otro lado, el modelo también tenía en cuenta la puntuación del HBS asignada por comparación con la matriz de frecuencias (PWM).

Para un gen dado existían dos estados posibles, ser diana de HIF o no serlo. Definimos una probabilidad asociada a cada estado y el gen se clasificaba de acuerdo al ratio (log ratio) de ambas probabilidades ( $P_t/P_b$ ).

Para establecer la ecuación que definía a un gen como diana de HIF ( $P_t$ ) tomamos un conjunto de genes regulados por HIF bien caracterizados y asignamos dos funciones de densidad de probabilidades, una asociada a la distribución de la magnitud de su modulación por hipoxia en los experimentos de microarrays  $f_t^f(x)$  y otra a la de las puntuaciones de sus HREs  $f_t^s(y)$ . De este modo, quedó la ecuación siguiente:

$$P(x, y, p|T) = f_t^s(\max(y_i)) * (f_t^f(x) + (p * f_b^f(x)))$$

Siendo  $f_t^s(\max(y_i))$  el valor máximo de HRE asociado a un locus y  $p$  el p-valor asociado a la modulación por hipoxia.

Para establecer la ecuación que definía a un gen como no-diana de HIF (Pb) se tomó como grupo de genes de referencia todos los genes del genoma, puesto que asumimos que la mayoría no se modulan por hipoxia. Así, esta ecuación se formuló como:

$$P(x, y, p|B) = f_b^s(\max(y_i)) * (f_b^f(x) + (p * f_t^f(x)))$$

El ratio de ambas ecuaciones representaba la probabilidad de ser o no un gen diana de HIF:

$$\frac{P_t}{P_b} = \frac{P(x, y, p|T)}{P(x, y, p|B)}$$

En las ecuaciones anteriores, la contribución relativa de la modulación de un gen por hipoxia, en la clasificación final, dependía del p-valor asociado a la media, de modo que si esta inducción no era significativa (cuando p-valor se aproxima a 1), la contribución de este parámetro en la ecuación final de probabilidad era similar en el caso de ser o no diana de HIF. Así, en este caso, la clasificación del gen venía dada solo por el valor asociado a su HRE, quedando el modelo probabilístico definido como:

$$\frac{P(x, y, p|T)}{P(x, y, p|B)} \approx \frac{f_t^s(\max(y_i))}{f_b^s(\max(y_i))}$$

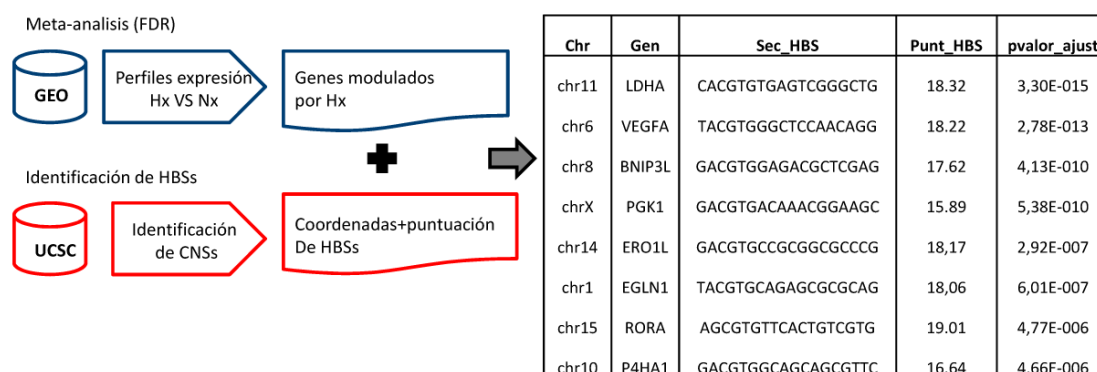
## ***RESULTADOS***



### V.1. ESTRATEGIA BIOINFORMÁTICA PARA LA IDENTIFICACIÓN DE NUEVAS DIANAS DE HIF.

El factor inducible por hipoxia, HIF, juega un papel central en la regulación transcripcional que media la respuesta a hipoxia, por ello, desde hace años, se ha invertido gran esfuerzo en tratar de identificar genes modulados por este factor. Recientemente se han llevado a cabo abordajes experimentales a gran escala, que permiten conocer globalmente la modulación transcripcional en condiciones de hipoxia (microarrays), así como la identificación de los sitios de unión de HIF en todo el genoma (Chip-Chip). A pesar de la gran información que aportan estas técnicas, también presentan limitaciones importantes como son: la variabilidad presente en los datos dependiendo del modelo celular y las condiciones experimentales utilizadas; y, en el caso de los microarrays, la imposibilidad de diferenciar dianas directas de HIF de indirectas o de dianas moduladas por otros factores de transcripción que se activen en hipoxia. Así mismo, ambas técnicas suponen una gran inversión de tiempo y dinero.

En este trabajo presentamos una estrategia bioinformática que evalúa el potencial de todos los genes del genoma como dianas de HIF. Esta estrategia combina datos públicos experimentales de expresión génica junto con la identificación *in silico* de sitios de unión de HIF en la totalidad del genoma (fig 12).



**Fig 12: Esquema de la estrategia usada para la clasificación de genes en diana/no-diana de HIF.** Ésta se basa en la información de los perfiles de expresión de GEO, con el que obtuvimos un patrón robusto de expresión génica en hipoxia, y en la identificación y puntuación de potenciales HBSs localizados en el genoma.

## Resultados

Los datos experimentales de perfiles de expresión fueron extraídos de la base de datos GEO (GEO, <http://www.ncbi.nlm.nih.gov/geo/>), y procesados utilizando un meta-análisis para identificar genes regulados por hipoxia. Para la identificación de los potenciales sitios de unión de HIF (HBSs) partimos de información genómica disponible en el servidor del UCSC y utilizamos una “huella filogenética” (phylogenetic footprinting) para identificar potenciales regiones reguladoras conteniendo HBSs, posteriormente le asignamos a cada HBS una puntuación que refleja la similitud de su secuencia con la de HREs funcionales. El resultado de la intersección de ambas informaciones permitió la clasificación de los genes es un listado fiel de genes diana/ no diana de HIF (fig 12).

### V.1.2. ANÁLISIS DEL PATRÓN DE REGULACIÓN DE LA EXPRESIÓN GÉNICA EN CÉLULAS SOMETIDAS A HIPOXIA.

Para intentar obtener un patrón robusto de expresión génica en condiciones de hipoxia, partimos de datos experimentales depositados en la base de datos pública Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>). Seleccionamos aquellos perfiles de expresión que comparaban el perfil de expresión de células en normoxia con el de células en hipoxia. Así extrajimos 249 perfiles de expresión, de los cuales, seleccionamos para nuestro estudio 16 de ellos en función de varios criterios:

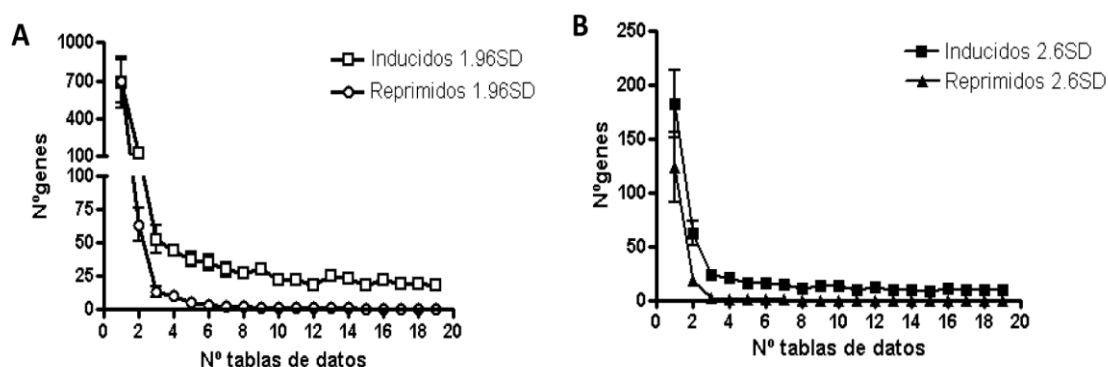
- Experimentos realizados en vertebrados, lo que permitía homogeneizar el conjunto de genes a estudiar.
- Experimentos realizados a tiempos cortos de hipoxia (6 a 24 horas de hipoxia), con lo que enriquecíamos los datos de partida en dianas directas de HIF y descartábamos la regulación génica consecuencia de adaptaciones a estados de hipoxia crónica.
- Experimentos con parámetros de calidad, que tuvieran al menos, duplicados de los datos y alguna referencia de calidad de la hibridación (p-valor).

En algunos de estos 16 experimentos las células se sometían a diferentes estímulos, como por ejemplo, hipoxia y DMOG, en estos casos, los datos fueron tratados como dos grupos independientes (a lo largo del texto nos referimos a todos los estímulos como hipoxia, a modo de simplificación). De modo que, finalmente, obtuvimos 19 tablas de datos provenientes de 16 experimentos microarrays independientes (ver materiales y métodos, tabla 5).

### V.1.2.1. Análisis individual de los perfiles de expresión génica.

Con el fin de estudiar la información contenida en los perfiles de expresión realizamos, en primer lugar, un análisis individual de estos perfiles. Así, para cada una de las tablas de datos calculamos el log-ratio Hipoxia/Normoxia de la señal media de fluorescencia de cada sonda (media de fluorescencia de sus réplicas), considerándola como inducida o reprimida cuando su log-ratio era mayor que 1.96 SD con respecto a la media de fluorescencia de ese conjunto de datos o, tomando un criterio más estricto, mayor que 2.6 SD. Consideramos que un gen estaba modulado cuando al menos una de sus sondas tenía un log-ratio por encima de estos valores.

A continuación buscamos un conjunto de genes que, de acuerdo con estos criterios, estuvieran modulados en hipoxia en **todos** los perfiles de expresión. La figura 13 representa el número de genes modulados en función del número de tablas de datos analizados, el resultado refleja que a medida que tomamos un número mayor de experimentos (tablas de datos), el número común de genes regulados por hipoxia en todas ellas va disminuyendo de forma exponencial. En el caso de los genes inducidos por hipoxia, veíamos que, por encima de 6 experimentos, el número de genes se estabilizaba, y encontramos 8 o 17 (según el punto de corte) genes inducidos de forma reproducible en todos los conjuntos de datos (tabla 6). La identidad de estos genes se muestra en la tabla 7. Por otro lado, no encontramos ningún conjunto de genes que se reprimiese por hipoxia de forma robusta cuando se comparaban más de cinco experimentos independientes.



**Fig 13: Genes modulados en los 19 conjuntos de datos analizados individualmente.** (A y B) Tomamos “n” (representado en los valores del eje de abscisas) tablas de forma aleatoria de entre las 19 disponibles y contamos el número de genes inducidos o reprimidos por hipoxia en todas las tablas de datos, tomando modulaciones por encima o por debajo de 1.96 SD (A) o 2.6 SD (B) con respecto a la media como significativas. El procedimiento se repitió diez veces para cada “n”, en las gráficas se representa el número medio de genes modulados por hipoxia en las diez repeticiones y las barras de error representan la SD.

19 experimentos GEO	PUNTO DE CORTE	
	2.6 SD	1.96 SD
Nºgenes inducidos	8	17
Nºgenes reprimidos	0	0
Nºtotal genes	11673	

**Tabla 6: Genes modulados en los 19 conjuntos de datos según el punto de corte escogido.** Tan solo 8 o 17 genes, de un total de 11673 que aparecen representados en todos los experimentos, se encuentran inducidos de forma significativa en los 19 experimentos de microarrays, tomando como puntos de corte 2.6 SD y 1.96 SD respectivamente, mientras que ninguno aparece reprimido de forma robusta.

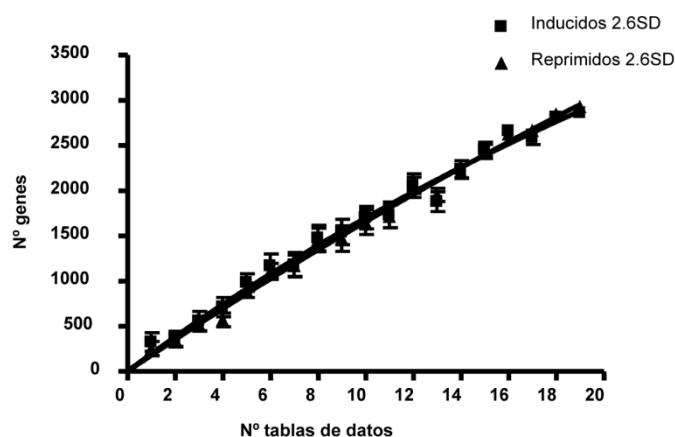
Símbolo Gen	Nombre del Gen	Inducción
VEGFA	Factor de crecimiento de endotelio vascular A	**
EGLN1	Hidroxilasa de HIF 2	**
PGK1	Fosfoglicerato kinasa 1	**
SLC2A3	Transportador de glucosa de membrana	**
BNIP3L	Proteína similar P 3 BCL2/adenovirus E1B	**
NDRG1	Proteína NDRG1 (Regulada por la ruta de N-myc)	**
HIST2H2BE	Histona H2B tipo 2-E	**
ERO1L	Proteína similar a ERO1	**
PDK1	Proteína kinasa1 dependiente de fosfoinositol	*
LDHA	Lactato deshidrogenasa A	*
AK3L1	Proteína1 similar a Adenilato kinasa3	*
BHLHB2	Proteína 40 basic helix-loop-helix (bHLHe40)	*
IGFBP5	Proteína de unión al factor de crecimiento IGF	*
BNIP3	19 kDa Proteína BCL2/adenovirus E1B	*
C3ORF28	Molécula proapoptótica de respuesta a HIF-1 $\alpha$	*
ZNF395	Proteína de dedos de zinc 395	*
CCNG2	Ciclina-G2	*

**Tabla 7: Genes inducidos en los 19 perfiles de expresión de GEO.** Del análisis individual de las 19 tabas de datos de GEO extrajimos un conjunto de genes que se encontraban inducidos en todos ellos. La tabla muestra la identidad de los genes cuya inducción era superior a 1.96 (\*) o 2.6 (\*\*) SD con respecto a la media.



Por otro lado, si estudiamos el número de genes no redundantes que aparecían modulados en los experimentos de microarrays, observamos que a medida que tomamos un número mayor de experimentos, el número de genes que aparecía modulado iba incrementando de forma casi lineal. De forma que, si tomábamos los 19 perfiles de expresión, y los analizábamos con el punto de corte más estricto (2.6 SD), encontramos que 2864 genes se encuentran inducidos y 2929 reprimidos (fig 14). Estos números suponían cerca del 50% (5793 genes de un total de 11673) de los genes representados en los 19 perfiles.

Por lo tanto, estos análisis nos indicaron que, la simple intersección de los datos experimentales (fig13) de los perfiles de expresión analizados individualmente era una aproximación demasiado restrictiva, en la que perdíamos gran cantidad de información (muchas dianas conocidas de HIF no están presentes en la tabla7). Por otro lado, la suma de información de estos análisis individuales (fig14) mostraba una cantidad de “ruido” (falsos positivos) excesiva, incluso tomando el punto de corte mas estricto, consecuencia de la suma del ruido de cada experimento individual.



**Fig 14: Conjunto de genes “no redundantes” modulados en los 19 experimentos de microarrays.** Representamos la media y desviación estándar de los genes registrados como inducidos o reprimidos (tomando 2.6 SD), para cada grupo de “n” experimentos. Vemos como a medida que se incrementa el número de experimentos tomados, crece de forma exponencial el número de genes que hallamos modulados.

#### V.1.2.2. Meta-análisis de los perfiles de expresión génica.

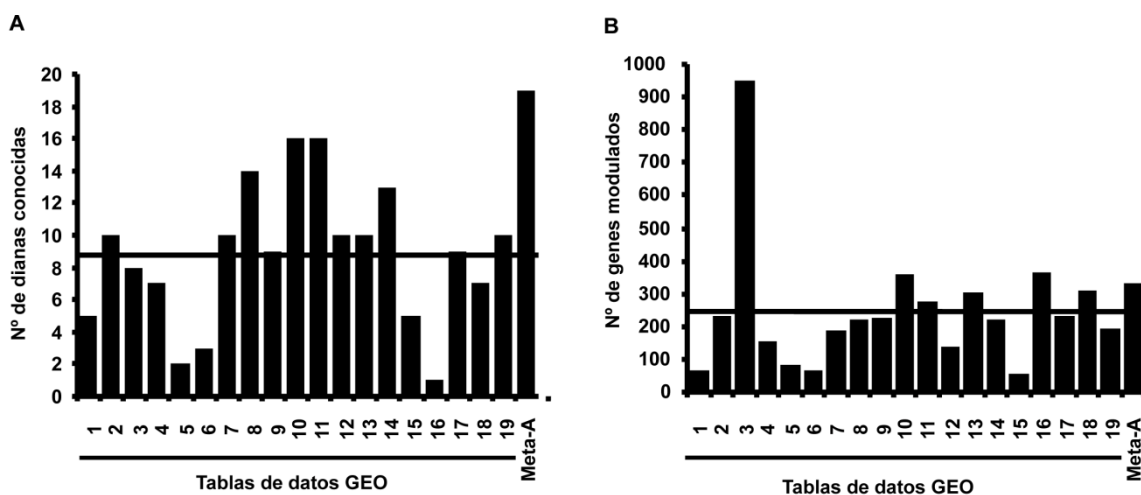
Con el fin de corregir las deficiencias de estos abordajes simplistas decidimos hacer un análisis combinado o meta-análisis. Para ello, en primer lugar, estandarizamos los valores de los log-ratio asociados a cada sonda para cada experimento independiente con el fin de poder comparar entre si las distintas tablas de datos. A continuación registramos los valores estandarizados de todas las sondas relativas a cada locus en todas las tablas de datos. Por

## Resultados

último, asumiendo una distribución normal para ese conjunto de datos, calculamos si la media de los log-ratio de todas las sondas relativas a un locus difería significativamente de cero (según la hipótesis nula de que el gen no se induzca por hipoxia, la media de sus log-ratio sería igual a cero) utilizando un t-test simple. Finalmente los p-valores obtenidos se corrigieron utilizando el método *FDR (False discovery rate)* y consideramos genes modulados por hipoxia aquellos que presentaban un p-valor corregido por debajo de 0.01 (ver materiales y métodos, IV.6).

Como resultado de este meta-análisis obtuvimos que 259 genes (2.22%) estaban inducidos por hipoxia y 191 genes (1.64%) estaban reprimidos por hipoxia, del total de los 11673 representados en todos los experimentos de GEO seleccionados.

Para evaluar la fiabilidad de este método de análisis, tomamos un grupo control de 65 genes considerados dianas reales de HIF (Wenger, Stiehl et al. 2005) y estudiamos cuantos de ellos aparecían modulados en cada uno de los 19 análisis individuales o en el meta-análisis. Como se puede ver en la figura 15, el meta-análisis era capaz de identificar un número mayor de dianas reales de HIF que cada uno de los análisis individuales. Así pues, concluimos que el meta-análisis tenía una mayor sensibilidad que cualquiera de los análisis individuales, sin reducirse la especificidad, puesto que el número total de genes identificados como modulados por hipoxia en el meta-análisis era similar al de los análisis individuales.



**Fig 15: Comparativa del meta-análisis frente al análisis individual de los datos de GEO.** (A) La gráfica representa el número de genes diana de HIF que aparecen como inducidos en cada uno de los análisis individuales de los perfiles de expresión de GEO (tomando como punto de corte una modulación mayor de 2.6 SD sobre la media) y en el meta-análisis (tomando como punto de corte un p-valor ajustado <0.01). (B) Número total de genes que aparecen como inducidos por hipoxia en ambos tipos de análisis (número de genes cuya modulación es mayor de 2.6 SD por encima de la media). La línea horizontal representa la media de los genes inducidos, diana y totales respectivamente, en los 19 experimentos de microarrays.

### V.1.3. IDENTIFICACIÓN IN SILICO DE SITIOS DE UNIÓN DE HIF EN LA TOTALIDAD DEL GENOMA.

El factor inducible por hipoxia HIF tiene un sitio de unión al genoma bastante estricto, RCGTC, sin embargo, debido a su pequeño tamaño se puede encontrar por azar en cada 512pb, lo que supondría que cada gen puede contener al menos un HBS. Por lo tanto es de esperar que exista un elevado número de motivos RCGTG no funcionales. Por ello, con el fin de intentar extraer solo sitios RCGTG funcionales, restringimos nuestro estudio a secuencias no codificantes conservadas evolutivamente o CNSs. Las regiones no codificantes conservadas evolutivamente son ricas en sitios de unión de factores de transcripción, por tanto esta restricción nos permitió incrementar la eficiencia de búsqueda de sitios de unión de HIF (HBS), así como reducir el número de sitios de RCGTG no funcionales. Además, esta asunción viene apoyada por el hecho de que, no se han descrito diferencias globales en la regulación de genes por HIF en vertebrados.

Para cada locus definimos sus CNSs localizando regiones PhasCons (Siepel, Bejerano et al. 2005), dentro de sus regiones no codificantes: regiones intrónicas, regiones 5' y 3'UTR y en la región promotora tomando 5kb por encima del sitio inicial de la transcripción. A continuación buscamos sitios RCGTG dentro de estas CNSs y, con el fin de reducir el ruido, solo consideramos como potenciales HBSs los motivos RCGTG conservados al menos en cuatro especies de vertebrados, incluyendo siempre ratón (materiales y métodos, fig 11). Con estas restricciones asumimos que la regulación mediada por HIF estaría conservada evolutivamente y por tanto, que para cada gen diana el HRE estaría presente en el ancestro común del grupo de vertebrados.

Como resultado de esta búsqueda, obtuvimos 9458 potenciales sitios de unión de HIF distribuidos en 3980 locus génicos, de modo que localizamos motivos RCGTG en el 34.1% de los genes analizados, mientras que en el restante 65.9% no encontramos ninguno de acuerdo con las restricciones impuestas.

## Resultados

### V.1.3.1. Asignación de un valor numérico a los HBSs localizados.

El alineamiento de 46 secuencias provenientes de 23 genes diana de HIF bien caracterizados (anexo II) reveló que, además del motivo RCGTG del HRE, existían bases en otras posiciones que se presentaban con mayor frecuencia que la esperada por azar. Con esta información establecimos una matriz (PWM, *Position Weight Matrix*) de 18 pb de longitud en la que se asignaba a cada base para cada posición, una frecuencia de aparición en la misma (ver materiales y métodos, IV.7).

Así, una vez que localizamos todos los motivos RCGTG del genoma (como se ha indicado en el apartado anterior), les asignamos una puntuación que definía su similitud de secuencia con la de HREs funcionales en base a la PWM definida.

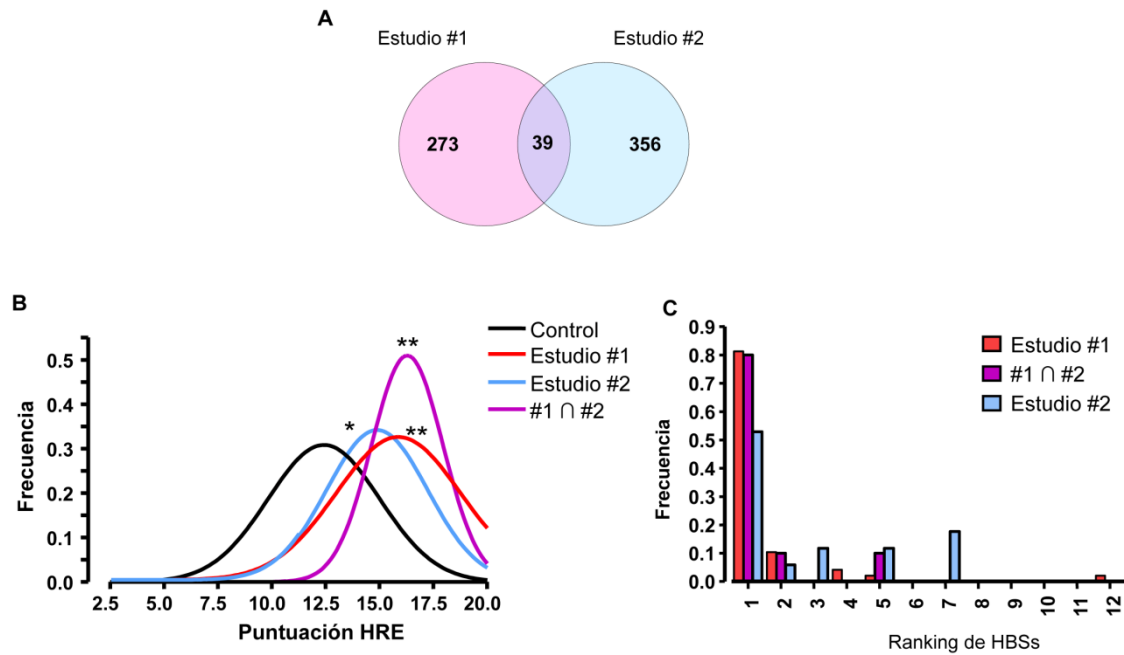
Para evaluar el valor de la puntuación derivada de la PWM a la hora de discriminar HREs funcionales, estudiamos la distribución de puntuaciones en los HBSs identificados en dos estudios experimentales de ChiP-ChiP para HIF $\alpha$  ((Mole, Blancher et al. 2009) y (Xia, Lemieux et al. 2009)) que se publicaron durante el desarrollo de este proyecto.

Entre ambos trabajos existía muy poco solapamiento en el conjunto de genes identificados como dianas de HIF (solo 39 regiones comunes aparecen en ambos trabajos) (tabla 8 y fig 16A), lo que puede venir dado por diferencias en las líneas celulares y las condiciones experimentales utilizadas (incluyendo diferentes plataformas de hibridación). Por ello, consideramos más fiables aquellos sitios de unión de HIF que aparecían en ambos trabajos.

La figura 16B muestra que la puntuación asignada, según nuestra PWM, a los HREs identificados experimentalmente (estudio#1 (Xia, Lemieux et al. 2009) y estudio#2 (Mole, Blancher et al. 2009)) era significativamente superior a la media de de todos los sitios RCGTG del genoma (control) y la diferencia era aún mayor para el conjunto de genes solapantes a ambos estudios (#1 $\cap$ #2). También es importante destacar que solo cuatro genes de los presentes en ambos trabajos se incluyeron en los alineamientos para generar la matriz, GAPDH, LDHA, PGK1 y TF, de modo que los resultados comparativos no quedarían sesgados por este hecho.

Símbolo gen	Chr	Trabajo#1 c.inicio	Trabajo#1 c.fin	Trabajo#2 c.inicio	Trabajo#2 c.fin
ALKBH5	17	18026863	18027796	18026917	18027548
ANKZF1	2	219803111	219803786	219803046	219803691
BCKDHA	19	46595105	46595845	46595493	46595714
BCL2L2	14	22840062	22840843	22839864	22840201
CP	3	150425395	150426332	150425412	150425864
CRKL	22	19600995	19601491	19600989	19601171
DARS	2	136458552	136459891	136458510	136459327
GAPDH	12	6512518	6514311	6513390	6513563
GBE1	3	81892424	81894007	81892043	81893145
GPI	19	39541621	39542821	39541340	39542126
HIG2	7	127882528	127883473	127882554	127883158
HK2	2	74914194	74914694	74914233	74914441
INHA	2	220148994	220150412	220149206	220150032
INSIG2	2	118561790	118562716	118561723	118562366
JMJD1A	2	86520885	86522113	86520791	86521639
KIAA0195	17	70960766	70961264	70960840	70961202
KIAA1715	2	176574095	176574949	176574234	176574603
LDHA	11	18372355	18373091	18372041	18372531
LOC401152	4	120440857	120441746	120440308	120441137
NARF	17	78008892	78009954	78008958	78009355
NCOA7	6	126143360	126144150	126143751	126143902
NFE2L1	17	43480925	43481728	43481265	43481686
PGK1	X	77245708	77246723	77245762	77245920
PLOD2	3	147361588	147362434	147362257	147362515
PLOD3	7	100647716	100648319	100647759	100647962
PNRC1	6	89847972	89848858	89848399	89848986
PPP1R3C	10	93382442	93383707	93383091	93383567
PSMA1	11	14498230	14498962	14498905	14499265
RAPGEF6	5	130998588	130998859	130998954	130999756
RIT1	1	154147622	154148413	154148103	154148331
RSBN1	1	114156314	114157000	114156633	114157374
SMEK2	2	55697825	55698799	55698553	55698820
SOS1	2	39204502	39205544	39204318	39204864
SPRY1	4	124537622	124538323	124537498	124538111
TF	3	134943955	134944566	134944094	134944856
TMEM45A	3	101693922	101694508	101694373	101695020
UGP2	2	63922930	63923673	63923104	63923497
WSB1	17	22644869	22645641	22644839	22645013
MTP18	22	29149472	29150541	29149767	29149951

**Tabla 8: Genes solapantes en los dos estudios de ChiP-ChiP.** En la tabla se presenta el símbolo correspondiente al gen identificado, el cromosoma en el que se localiza, y las coordenadas correspondientes a la región inmunoprecipitada en cada trabajo.



**Fig 16: Correlación entre una elevada puntuación del HRE y la funcionalidad del mismo.** (estudio#1(Xia, Lemieux et al. 2009) y estudio#2(Mole, Blancher et al. 2009)) (A)Diagrama de Venn que muestra la intersección de los genes identificados los dos trabajos de ChiP-ChiP. (B) La distribución de las puntuaciones asignadas a cada grupo de HBSs fue ajustada a una distribución Gaussiana. La gráfica muestra las distribuciones de puntuaciones de los HBSs identificados en cada uno de los trabajos de ChiP-ChiP (trabajo#1 y trabajo#2), las puntuaciones de los genes comunes a ambos (#1∩#2) y las de un grupo control (motivos RCGTG localizados en todas las CNSs del genoma). Los valores medios de puntuación de cada grupo se compararon mediante un test de ANOVA. Las diferencias estadísticamente significativas con el grupo control están marcadas con asterisco (\*,  $p < 0.01$ ; \*\*,  $p < 0.001$ ). (C) Los motivos RCGTG identificados en un locus dado se ordenaron de mayor a menor según la puntuación dada por la PWM, a continuación se registró la posición (rank) del motivo que coincide con alguno de los identificados experimentalmente. La gráfica muestra la fracción de los motivos identificados *in silico* coincidentes con los experimentales ordenados en un ranking de puntuaciones.

Puesto que para algunos locus nuestra estrategia localizaba más de un HBS, quisimos estudiar si el valor de puntuación asignado a cada sitio era suficiente para identificar motivos RCGTG funcionales de un locus. Para ello tomamos genes con varios sitios HBSs identificados por nuestra estrategia, y analizamos cual de los HBSs (ordenados en un ranking de mayor a menor puntuación asignado a su HBS) coincidía con un HBS identificado experimentalmente en los estudios de Chip-Chip (fig 16C). Este análisis demostró que de todos los HBSs predichos para un gen dado, el de mayor puntuación (posición más alta en el ranking) coincidía, en la mayoría de los casos, con el HBS identificado experimentalmente.

#### V.1.4. MODELO PROBABILÍSTICO QUE INTEGRA DATOS DE EXPRESIÓN GÉNICA E IDENTIFICACIÓN DE HBSs.

En base a los resultados anteriores, teníamos asociado a cada uno de los genes del genoma dos informaciones (anexo IV):

- Información sobre su modulación por hipoxia en base a los resultados del meta-análisis. Para cada locus génico teníamos la magnitud de su modulación por hipoxia ( $\log Hx/Nx$ ) y el p-valor ajustado asociado a esta modulación que indica como de significativa es.
- Información sobre los potenciales sitios de unión de HIF asociados a su locus, concretamente las coordenadas genómicas del sitio HBS y su puntuación de acuerdo con la PWM.

A partir de estas dos informaciones estudiamos la correlación entre la presencia de HBS y la modulación por hipoxia. Así, encontramos que el número de genes que presentaban al menos un HBS (con independencia de su puntuación) y se inducían por hipoxia ( $\log\text{-ratio}>0$  y  $p<0.01$ ) era mayor que el esperado por azar (test Chi-cuadrado, p valor=  $3.18 \times 10^{-14}$ ) mientras que esto no se cumplía en el caso de los genes reprimidos (p-valor= 0.42) (tabla 9). Concluimos que, en términos globales, HIF media la inducción de genes en respuesta a hipoxia y no así su represión, de hecho, tradicionalmente, la respuesta a hipoxia mediada por HIF se ha asociado mayoritariamente con una inducción de la expresión génica, siendo muy minoritarios los casos de represión transcripcional (Chen, Lai et al. 2005; Wen, Ding et al. 2010).

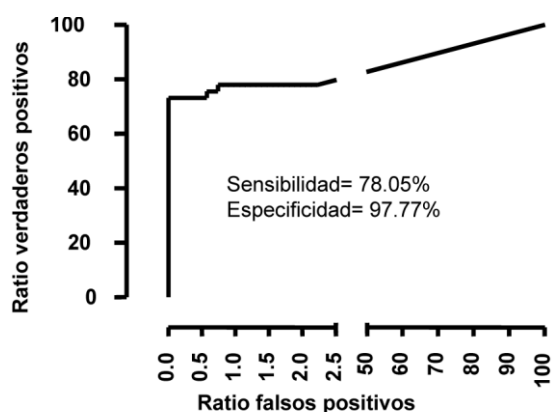
	HBS+		HBS-		p-valor
	Observados	Esperados	Observados	Esperados	
<b>Inducidos</b>	145	88	113	170	3.18x10 <sup>-14</sup>
<b>No-regulados</b>	3766	3823	7459	7402	
<b>Reprimidos</b>	69	64	121	126	0.42
<b>No-regulados</b>	3766	3771	7459	7454	

**Tabla 9: La presencia de HBS correlaciona con la inducción pero no con la represión transcripcional.** Los genes se clasificaron como inducidos, reprimidos y no regulados según su  $\log\text{-ratio}$  derivado de los perfiles de expresión de GEO y su p-valor ajustado asociado. El número de genes que presentan o no HBS (HBS+ y HBS-) es mostrado en Observados. El número de los esperados por azar se muestra en las columnas de esperados. La significancia entre los valores observados y esperados se calculó con un test de Chi-cuadrado y el resultado del test queda reflejado en el parámetro del p-valor.

## Resultados

En base a este resultado, en el resto de nuestro estudio nos centramos solamente en genes inducidos por hipoxia.

Para identificar genes inducidos por HIF, asumimos que cada gen podía estar en uno de estos dos estados: ser diana de HIF o no serlo. Así, calculamos dos funciones de probabilidad asociadas a cada gen de la tabla, la probabilidad de ser diana de HIF ( $P_t$ ) frente a la probabilidad de no serlo ( $P_b$ ) (es a lo que denominamos background). Finalmente el valor del cociente  $P_t/P_b$  era el que clasificaba al gen como diana o no de HIF (ver materiales y métodos). Para poder clasificar los genes en función del parámetro  $P_t/P_b$  era necesario establecer un punto de corte que nos permitiera detectar genes diana de HIF con alta sensibilidad pero sin perder la especificidad. Para ello realizamos un análisis de curvas ROC (*receiver operating characteristic*) (fig 17) en el que utilizamos dos grupos control: el grupo control positivo, que era el conjunto de 39 genes comunes en los dos ensayos de ChiP-ChiP (tabla 8) como verdaderas dianas de HIF; y como grupo de control negativo utilizamos 1221 genes que, a pesar de tener HBSs asociados, no aparecían como modulados en los 19 perfiles de expresión génica (valor de inducción entre -0.3 y +0.3 y un  $p$ -valor > 0.5) (anexo III). En base a este análisis establecimos como punto de corte un valor de  $P_t/P_b > 6.5$  a partir del cual consideramos los genes como dianas de HIF. Este punto de corte muestra que el modelo de clasificación tiene una sensibilidad del 78.05% y una especificidad del 97.77%.

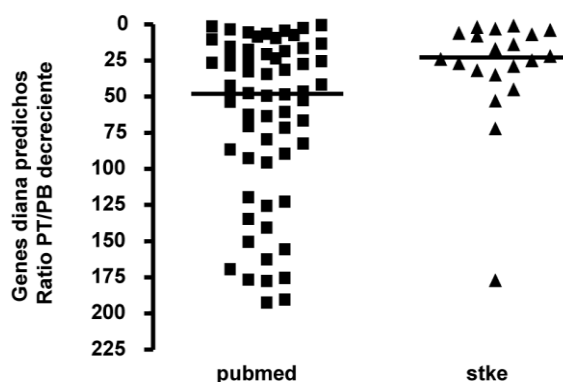


**Fig 17: Determinación del valor de  $P_t/P_b$  como punto de corte mediante curva ROC (*Receiver Operating Characteristic*).** Para este análisis usamos dos grupos de genes control para los que calculamos su valor de  $P_t/P_b$ : 39 genes diana de HIF comunes a los dos estudios de Chip-Chip como verdaderos positivos, y 1221 genes que contenían HBSs pero no se modulaban por hipoxia como negativos. El punto de corte del ratio  $P_t/P_b$  que maximiza la sensibilidad y especificidad para este conjunto de datos es 6.5.



Con este modelo clasificamos los 11673 genes para los que disponíamos información relativa a su modulación por hipoxia (aparecían en el meta-análisis). Obtuvimos un listado final en el que 216 genes se clasificaron como dianas de HIF, de los cuales, 20 estaban descritos previamente (Xia, Lemieux et al. 2009), de otros 44 existía información bibliográfica que los relacionaba con la respuesta a hipoxia y los 152 restantes son nuevas potenciales dianas de HIF (anexo IV).

Para analizar el valor del cociente Pt/Pb como predictor de dianas de HIF representamos los 20 genes previamente descritos (Wenger, Stiehl et al. 2005) y los 44 que se han relacionado con la hipoxia ordenados según su valor de Pt/Pb. Como muestra la figura 18 todos ellos se posicionan en torno a los valores más altos de Pt/Pb (primeros puestos del ranking), de modo que este parámetro parece ser buen indicador de genes diana de HIF reales.



**Fig 18: El valor Pt/Pb parece ser un buen indicador de genes diana de HIF.** En la gráfica mostramos los valores Pt/Pb asociados a los 20 genes diana de HIF bien caracterizados (stke, (Wenger, Stiehl et al. 2005)) y los 44 de los que había información bibliográfica relacionándolos con la respuesta a hipoxia (pubmed). Todos ellos se distribuyen mayoritariamente en torno a los valores más altos de Pt/Pb (posicionados en los primeros puestos del ranking).

## V.2. VALIDACIÓN EXPERIMENTAL DE GENES DIANA DE HIF PREDICHOS POR NUESTRA ESTRATEGIA BIOINFORMÁTICA.

Como resultado de la aplicación de nuestra estrategia bioinformática obtuvimos un listado final de 3980 genes que contenían uno o varios sitios RCGTG, de los que 216 aparecían clasificados como potenciales dianas de HIF (anexo IV). Para 152 de estos genes no existía ninguna evidencia bibliográfica previa de su relación con la respuesta a hipoxia.

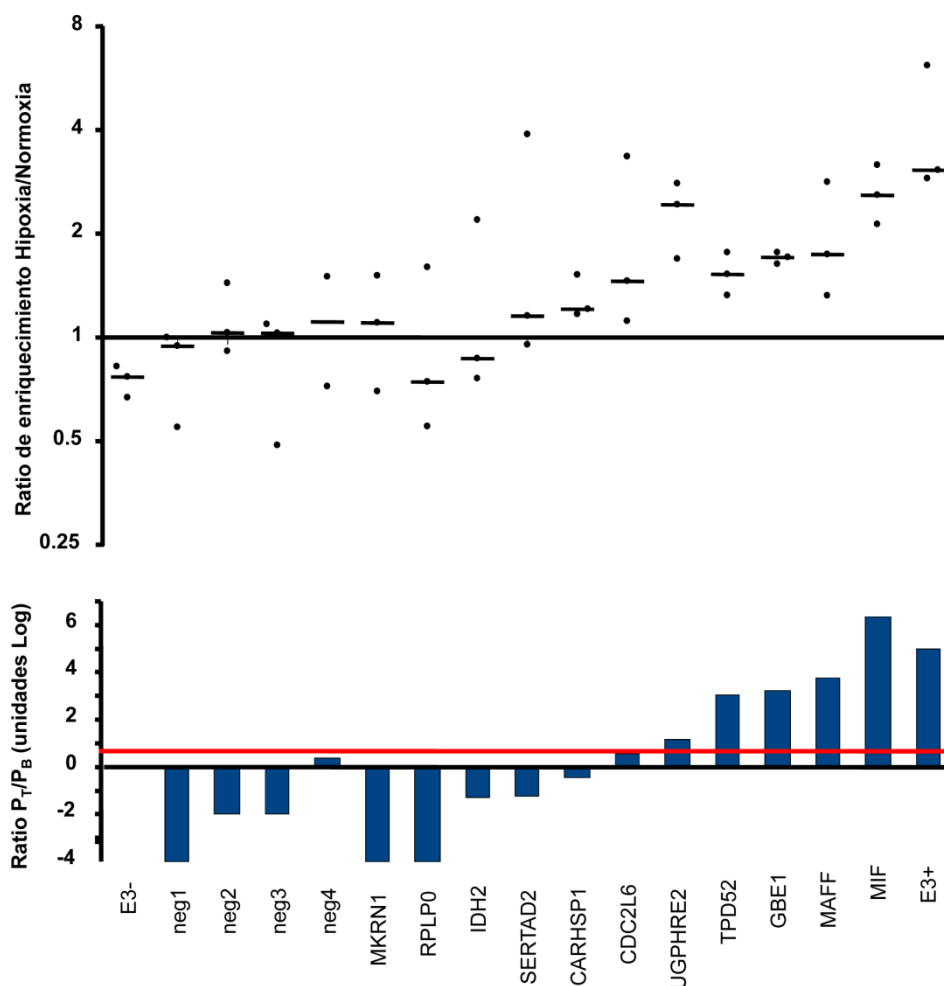
## Resultados

Con el fin de validar la unión de HIF a los HBSs predichos, realizamos ensayos de Chip para HIF1 $\alpha$ . Seleccionamos al azar once genes del listado total de genes inducidos, según el meta-análisis, que presentaban un solo sitio de unión de HIF (para simplificar el análisis). Analizamos la unión de HIF1 $\alpha$  a los HBSs identificados, mediante ChIP y qPCR-RT. De estos once genes, seis aparecían clasificados como dianas de HIF (CDC2L6, UGP2, TPD52, GBE1, MAFF y MIF) y los otros cinco (MKRN1, RPLP0, IDH2, SERTAD2 y CARSH1), como no-dianas de HIF, según el punto de corte establecido (Pt/Pb=6.5). También seleccionamos cuatro genes como controles negativos, con el fin de establecer el nivel basal de enriquecimiento hipoxia/normoxia (background), los cuales, presentaban un sitio RCGTG asociado, pero no se modulaban por hipoxia de acuerdo con nuestro meta-análisis (neg1= IRS4, neg2= STT3A, neg3= HIVEP y neg4= LTBP1). Así mismo, tomamos los dos HREs asociados al gen de la EglN3, de los cuales solo uno de ellos es funcional, (Pescador, Cuevas et al. 2005) como controles de sitio real de unión de HIF y un sitio RCGTG no funcional para la unión.

Los ensayos de Chip (*Chromatin Immunoprecipitation*) se realizaron en células Hela sometidas a condiciones de normoxia o hipoxia durante 6 horas, su DNA fue inmunoprecipitado usando un anticuerpo específico de HIF1 $\alpha$  y la unión de HIF1 $\alpha$  a los HBSs se determinó por qPCR-RT en las muestras del inmunoprecipitado, finalmente calculamos los ratios de enriquecimiento de los inmunoprecipitados en hipoxia frente a normoxia, mostrando estos valores en porcentaje con respecto al material de partida o input (ver materiales y métodos).

En la figura 19 se muestran los resultados de las validaciones experimentales y las predicciones bioinformáticas (ratio Pt/Pb) para cada gen. En las tres réplicas biológicas observamos que el ratio de enriquecimiento de los controles negativos era cercano a uno (background, log Pt/Pb), como era de esperar, la hipoxia no aumenta la unión de HIF a estas regiones en comparación con la normoxia. Por otro lado la unión de HIF al sitio usado como control positivo, el HRE funcional de EglN3 (E3+), tenía un enriquecimiento elevado. De los genes inducidos por hipoxia según el meta-análisis vimos que, solo aquellos con un valor de Pt/Pb>6.5 eran los que realmente unían HIF en los tres experimentos independientes (UGP2, TPD52, GBE1, MAFF1 y MIF), y a pesar de la dispersión que presentaban los datos experimentales, existía una buena correlación entre validación experimental y predicción bioinformática. Lo mismo se cumplía para los genes predichos como no diana de HIF (Pt/Pb<6,5), los cuales presentaban un patrón muy similar al de los controles negativos. Existían dos excepciones, CAHSP1 y CDCL6, cuyos datos experimentales no se ajustaban a la predicción, posiblemente porque para ambos casos, sus ratios Pb/Pt son muy próximos al

punto de corte ( $Pb/Pt=6.5$ ). Por tanto, nuestra validación experimental puso de manifiesto que cinco de seis genes clasificados como dianas de HIF lo eran y que cuatro de cinco clasificados como no-dianas eran verdaderos negativos.

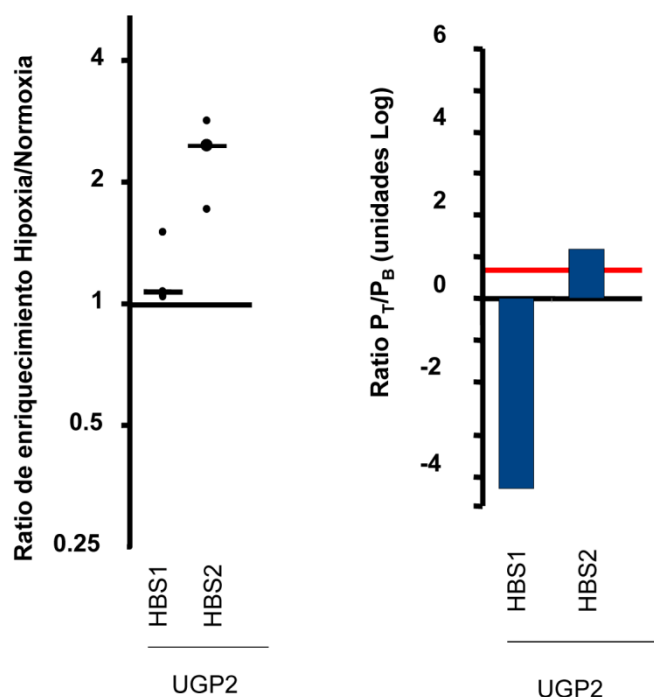


**Fig 19: Validación experimental de los HBSs predichos por nuestra estrategia bioinformática.** Para las validaciones se realizaron tres réplicas biológicas de experimentos de inmunoprecipitación de cromatina para HIF1 $\alpha$  en células Hela expuestas a condiciones de normoxia/hipoxia. La identificación de la unión de HIF a las regiones predichas se realizó por qPCR-RT. En la parte inferior de la figura se muestran los valores de los ratios Pt/Pb asociados a cada gen y la línea roja es el punto de corte para el valor Pt/Pb ( $\log_{10} 6.5$ ). En la parte superior se muestra el porcentaje de enriquecimiento de los inmunoprecipitados hipoxia/normoxia de las tres réplicas biológicas. Los puntos negros representan los valores de enriquecimiento en cada uno de los experimentos individuales y la línea negra la mediana de los tres experimentos.

Como prueba adicional de la importancia del valor asignado al HBS en las predicciones, estudiamos la unión de HIF a los dos sitios RCGTG identificados en el gen UGP2. Este gen se induce por hipoxia según el meta-análisis, y sus dos HBSs conservados evolutivamente estaban localizados en el cromosoma 2 en las posiciones 63922445 y 63923300. Los valores de

## Resultados

Pt/Pb asociados a estos potenciales HBSs eran  $3 \times 10^{-5}$  y de 14.4 respectivamente, en los que el valor de puntuación de sus HBSs (10.11 y 14.81) determinaba la clasificación del gen en diana o no. Así, analizamos la unión de HIF1 $\alpha$  a estos dos HBS y tal y como se muestra en la figura 20, vimos como la predicción para ambos HBSs correlacionaba con la validación experimental. Con este resultado se pone de manifiesto, una vez más, el valor predictivo de la puntuación asignada al HBS a la hora de clasificar a un gen como diana o no de HIF.



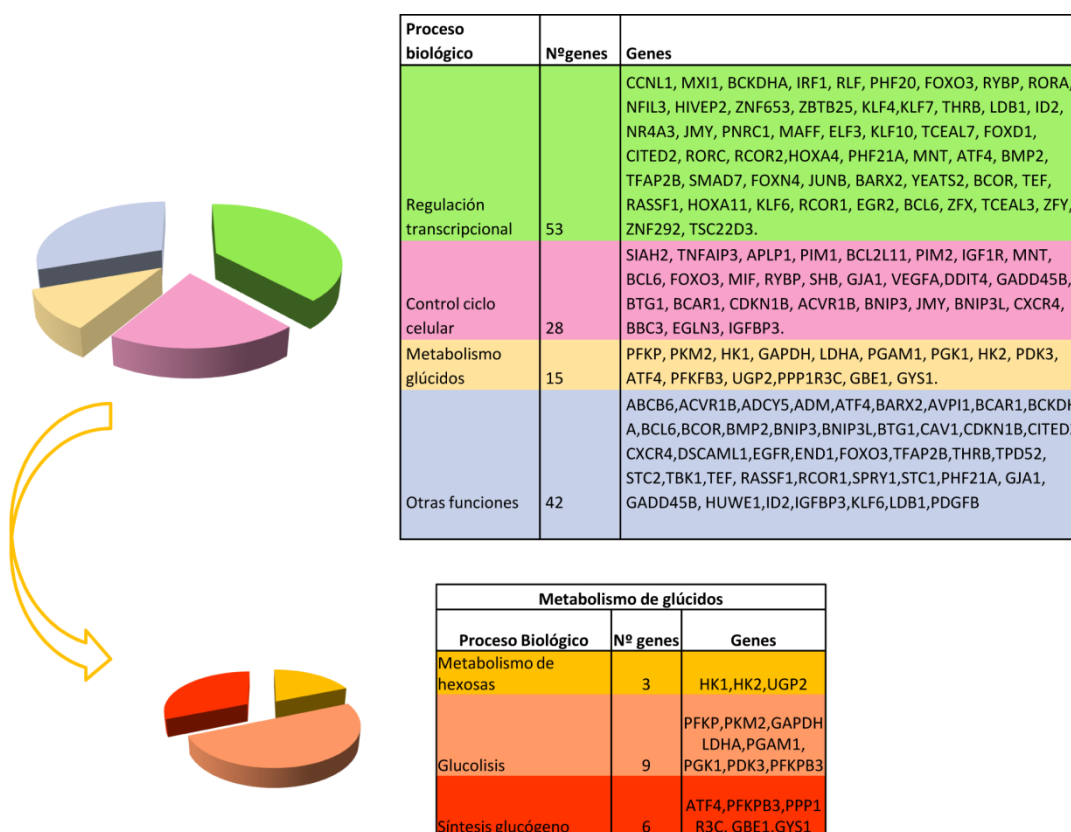
**Fig 20: Validación de los dos HBSs asociados al gen UGP2.** Como en la figura anterior, el gráfico de la derecha muestra los valores Pb/Pt asociados a cada HBS, la línea roja es el background. El gráfico de la izquierda muestra los valores de % con respecto al input de enriquecimiento hipoxia/normoxia de los inmunoprecipitados, se muestran las tres réplicas biológicas (puntos negros) y la mediana (línea negra) para los dos HBSs.

Atendiendo a las validaciones experimentales podemos concluir que, la estrategia diseñada es capaz de clasificar genes diana de HIF con una alta sensibilidad y especificidad, obteniendo así un listado fiable de potenciales dianas de HIF.

### V.3. ESTUDIO FUNCIONAL DE LAS PREDICCIONES BIOINFORMÁTICAS.

Con el fin de conocer de forma global la respuesta a hipoxia mediada por HIF, estudiamos las funciones biológicas representadas por el conjunto de genes diana

identificados por la estrategia bioinformática. Para ello utilizamos el algoritmo de DAVID (DAVID, <http://david.abcc.ncifcrf.gov/>), que permite extraer las funciones biológicas más representativas de un listado de genes dado, englobando, en un mismo grupo funcional, a aquellos que presenten funciones redundantes o similares (Huang da, Sherman et al. 2009). Para ello, la aplicación se basa en la información contenida en diversas bases de datos, de todas ellas utilizamos para este análisis las anotaciones de ontología (GO, Gene Ontology). Esta anotación permite clasificar los genes en tres diferentes categorías atendiendo al proceso biológico en el que participan, componente celular al que pertenecen y su función molecular. Así, clasificamos el listado de genes diana según el proceso biológico en el que estuvieran implicados (escogiendo la clasificación GOTERMBP\_ALL). El resultado de este análisis (anexo V) puso de manifiesto que, en respuesta a hipoxia, las funciones biológicas representadas por un mayor número de genes diana de HIF eran las implicadas en regulación transcripcional, control del ciclo celular, y metabolismo de glúcidos principalmente (fig 21).



**Fig 21: Distribución de las funciones biológicas representadas por los genes diana de HIF en respuesta a hipoxia según su anotación en Gene Ontology (DAVID, <http://david.abcc.ncifcrf.gov/>).** De los 216 genes del listado, la aplicación clasificó 138, en la figura se muestran solo los grupos más representativos. Dentro del grupo funcional del metabolismo de glúcidos podemos establecer tres subcategorías: metabolismo de hexosas, glucolisis y síntesis de glucógeno.

V.3.1. NUEVOS GENES DIANA DE HIF IMPLICADOS EN CAMBIOS EN EL METABOLISMO DE LA GLUCOSA EN CONDICIONES DE HIPOXIA.

Un análisis más detallado de los genes implicados en el metabolismo de glúcidos reveló que, además de los efectos conocidos sobre la glucólisis, el metabolismo del glucógeno estaba sobrerrepresentado. Aunque en algún estudio previo ya se había mostrado acumulo de glucógeno en condiciones de hipoxia (Mamedova, Shneyvays et al. 2003; Vigoda, Mamedova et al. 2003), no se había descrito el mecanismo molecular subyacente, y el metabolismo del glucógeno no se había considerado hasta ahora como uno de los procesos regulados por HIF. Nuestro estudio reveló la existencia de cuatro genes diana de HIF que controlan en puntos clave la ruta de síntesis de glucógeno: GYS1 (isoforma muscular de la glucógeno sintasa), UGP2 (UTP glucosa 1 fosfato uridil transferasa), GBE1 (enzima ramificadora de glucógeno 1-4) y PPP1R3C (subunidad de la proteína PP1, activadora de la GYS1) (fig 22).

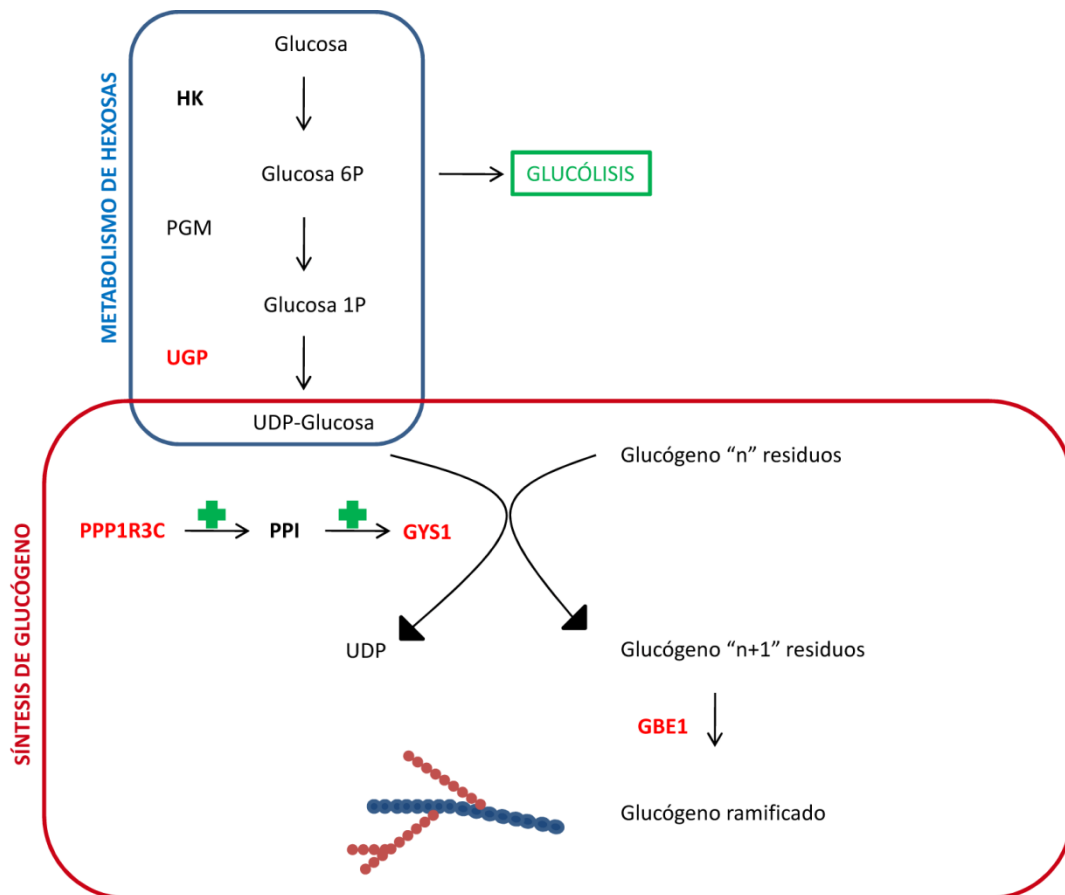
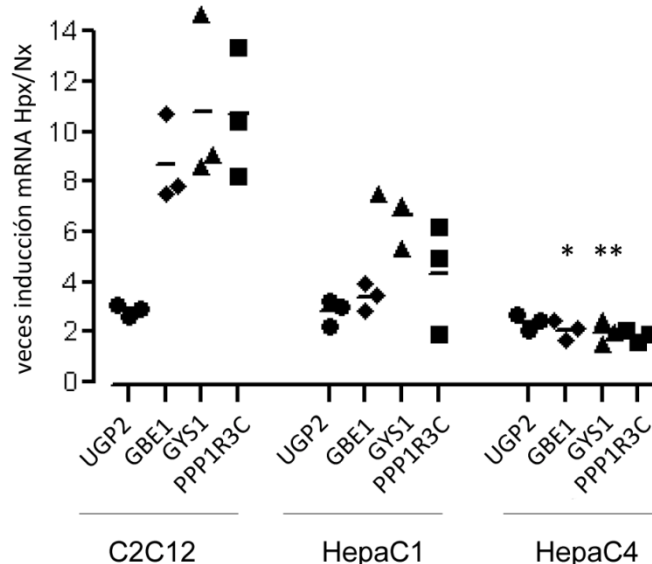


Fig 22: Procesos del metabolismo de la glucosa en los que están implicados nuevos genes diana de HIF. En rojo aparecen marcadas las enzimas moduladas por HIF, de las cuales UGP2, PPP1R3C, GYS1 y GBE1 son nuevas dianas anteriormente no descritas.

Con el fin de verificar esta nueva función metabólica de la hipoxia, puesta de manifiesto en el análisis de ontología, estudiamos la inducción por hipoxia de los nuevos genes diana implicados en la ruta de síntesis de glucógeno. Como modelo de estudio tomamos la línea muscular c2c12, ya que el músculo esquelético es uno de los principales reservorios de glucógeno.



**Fig 23: Regulación transcripcional de genes implicados en metabolismo de síntesis de glucógeno de manera HIF dependiente.** Las células c2c12 y Hepac1/c4 se expusieron a condiciones de normoxia o hipoxia durante 12 horas. Los mensajeros de los genes diana fueron cuantificados por qPCR-RT. Las gráficas muestran los ratios de inducción de hipoxia/normoxia, los símbolos negros representan los tres experimentos independientes y la línea negra la media de los mismos. GYS1, UGP2, GBE1 y PPP1R3C se inducen en hipoxia en células c2c12 y HepaC1, mientras que en HepaC4 los valores de expresión de estos genes son similares en hipoxia y en normoxia. (\*  $p < 0.05$ , \*\*  $p < 0.01$ )

Las células c2c12 se diferenciaron a células musculares y se expusieron a condiciones de normoxia o hipoxia durante 12 horas, tras las cuales extrajimos su RNA y cuantificamos por qPCR-RT los niveles de mensajero de los nuevos genes diana implicados en esta ruta (UGP2, GYS1, PPP1R3C y GBE1). Como se puede observar en la figura 23 estas cuatro enzimas ven aumentada su expresión en hipoxia en las tres réplicas biológicas independientes. Por otro lado, con el fin de relacionar directamente su modulación con HIF, hicimos este mismo ensayo en células HepaC1 y HepaC4, línea salvaje y mutante deficiente en la subunidad HIF $\beta$  respectivamente. Tal y como se muestra en la figura 23, las inducciones de estos enzimas se mantenían en HepaC1, pero se perdían totalmente en HepaC4, lo que implicaba directamente a HIF en su modulación transcripcional en hipoxia.

En conjunto, todos estos resultados validan nuestra estrategia para la identificación de genes diana funcionales y, además, sugieren una nueva función biológica hasta ahora no descrita como regulada por HIF en condiciones de hipoxia.

#### **V.4. IDENTIFICACIÓN DE SNPs QUE SE LOCALIZAN EN SITIOS DE UNIÓN DE HIF.**

HIF juega un papel central en la regulación transcripcional en respuesta a hipoxia, permitiendo la adaptación de la célula a esta bajada de oxígeno. Esta respuesta está mediada por la unión de HIF a los HREs de sus genes diana promoviendo su transcripción, por tanto, la presencia de cambios en la secuencia del núcleo central del HRE puede llevar a la pérdida de unión, y a una desregulación de estos genes en respuesta a hipoxia.

Los SNPs, polimorfismos de único nucleótido, son variaciones en la secuencia de DNA que afectan a una sola base y que, en la mayor parte de los casos, se trata de cambios puntuales de sustitución de una citosina por una timina. Son las responsables de cerca del 90% de las variaciones genómicas humanas y, por tanto, responsables de la variabilidad génica entre individuos, así como de determinadas enfermedades, respuestas alteradas a tratamientos farmacológicos...etc.

La presencia de SNPs en la secuencia del HRE podría generar una respuesta alterada a hipoxia, como consecuencia de la pérdida de unión de HIF a la misma. Esta desregulación podría llevar a incrementar la susceptibilidad de un individuo a ciertas enfermedades, o a una peor prognosis de un tratamiento en enfermedades que cursan con hipoxia. Por ello decidimos tratar de identificar los posibles SNPs que mapean en sitios HRE.

Para ello, buscamos SNPs que se localizasen dentro de las coordenadas genómicas de los motivos RCGTG de los genes diana obtenidos en la aplicación de la estrategia bioinformática (anexo VI). Las coordenadas de los SNPs las extrajimos de la base de datos dbSNP del servidor del NCBI ([www.ncbi.nlm.nih.gov/projects/SNP/](http://www.ncbi.nlm.nih.gov/projects/SNP/)) y obtuvimos un listado de SNPs que se localizaban en HBSs (anexo VI). De ellos, 12 correspondían a HBSs de genes clasificados como potenciales dianas de HIF (tabla 10).

Primeramente decidimos validar la unión de HIF a los HBSs de los genes MIF y CDC2L6 en los que localizaban los SNPs rs17004038 y rs10624, puesto que eran genes que tenían asociado un único HBS, lo que simplificaba la validación. Como se puede observar en la figura 18, ambos HBS unían HIF en ensayo de ChiP. Posteriormente, nos centramos en el SNP que se



localizaba en el HRE de MIF con el fin de estudiar la funcionalidad de su HRE y ver si la presencia de una mutación igual a la del SNP realmente alteraba la inducción de MIF por hipoxia.

SNP_ID	Base datos. Validación	Alelo	SNP_HBS	Gen	Chr	HBS_Loc	Max_HBS
rs13358075	hapmap	T/A	ACG_G	SPOCK1	5	136374995	NO
rs2901215	NA	T/G	ACG_G	TIAL1	10	121346556	NO
rs56306258	NA	A/T	_CGTG	ANKRD12	18	9126754	NO
rs56033752	NA	G/A	AC_TG	DLG2	11	83576587	NO
rs16943318	cluster,freq,hapmap	G/A	CAC_C	RORA	15	58997259	NO
rs3758554	freq	C/G	CA_GC	LDB1	10	103864659	NO
rs34394782	NA	C/A/T	CACG_	CITED2	6	139737161	NO
rs34476700	NA	G/-	_CGTG	SENP3	17	7404132	NO
rs56298217	NA	G/A	GC_TG	CLK3	15	72701699	NO
rs17152486	freq,hapmap	C/T	A_GTG	FLJ23834	7	105459125	SI
rs58433430	NA	C/T	A_GTG	KLHL14	18	28519311	SI
rs17004038	cluster,freq	C/A	_ACGT	MIF	22	22566590	SI

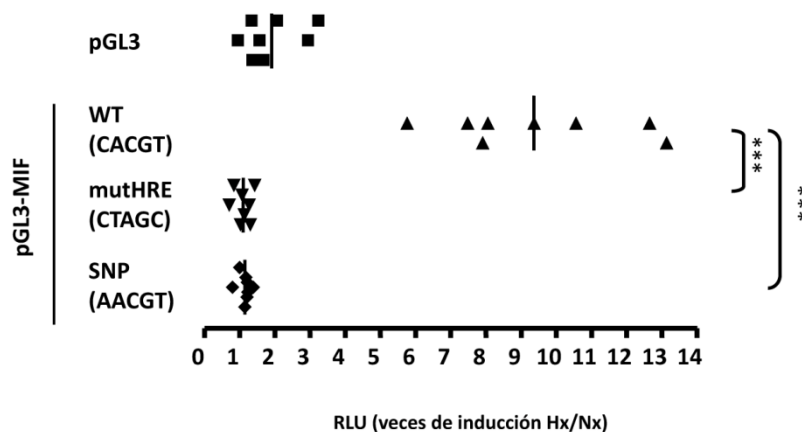
**Tabla 10: SNPs que se localizan en sitios HBSs de genes clasificados como diana de HIF.** En la tabla se muestra el número identificativo del SNP, la base de datos de donde proviene, el posible cambio de base que se produce en el alelo, la posición del polimorfismo en el motivo RCGTG, el gen al que se asocia el SNP\_HBS, el cromosoma en el que se encuentra y las coordenadas genómicas en las que localizan los HBS y finalmente, en el caso de que el gen tenga asociado más de un HBS si la modificación afecta al de mayor puntuación del locus.

Así, clonamos en el vector reportero *pGL3 Basic-luciferasa* la secuencia nativa de su promotor comprendida entre las coordenadas 22566534 y 22566673 del cromosoma 22 (Baugh, Gantier et al. 2006), que contenía el HRE. A partir de ella generamos dos construcciones mutantes, la primera eliminaba totalmente la funcionalidad del HRE (Chr22: 22566590) que cambiaba su secuencia de CACGT en la construcción salvaje a CTAGC en la mutante; la segunda introducía una mutación que mimetizaba el SNP que altera el motivo RCGTG con un cambio de la secuencia salvaje a AACGT.

Estas tres construcciones y el vector vacío de *pGL3Basic-luc* se transfectaron junto con un plásmido que codificaba para el enzima renila en células Hela y posteriormente fueron expuestas durante 12 horas a normoxia o hipoxia. Finalmente se midió la actividad renila-luciferasa por duplicado para cada transfección y los datos se procesaron corrigiendo con respecto a la renila, como normalización de la eficiencia de transfección. Se realizaron ocho

## Resultados

réplicas biológicas independientes de este experimento. Como podemos ver en la figura 24, la construcción original, WT, tenía un ratio de actividad luciferasa corregida en hipoxia frente a normoxia de una media de 9,5 veces de inducción, mientras que la construcción mutante mutHRE perdía totalmente la inducción por hipoxia. Estos dos datos en conjunto demostraron la funcionalidad del HRE predicho por nuestra estrategia bioinformática. En el caso de la construcción mutante SNP, observamos que esta mutación producía el mismo efecto que la anterior, reduciéndose la inducción de la actividad luciferasa en hipoxia frente a normoxia. Concluimos por tanto, que la presencia de este cambio impide la regulación transcripcional de MIF por parte de HIF, probablemente debido a una pérdida de unión del factor a su secuencia HRE.



**Fig 24: La mutación C/A introducida por el SNP rs 17004038 impide la inducción de MIF por hipoxia.** Las células Hela se transfectaron con las construcciones; wt, que comprendía la región promotora -31 a +108 delante del gen de la luciferasa; mut HRE, en la que se mutó su HRE (CACGT) a CTAGC y SNP, en la que se mutó a AACGT. Las células se sometieron a condiciones de normoxia o hipoxia y se midió su actividad luciferasa. La gráfica representa los ratios de actividad luciferasa hipoxia/normoxia. Los datos corresponden a ocho experimentos independientes y sus medias están marcadas como una línea vertical. Las diferencias significativas con respecto al grupo control son marcadas con asteriscos (\*  $p < 0.001$ ).

En conjunto, estos resultados suponen una prueba de concepto de la hipótesis de que la presencia de alteraciones en el núcleo central de la secuencia de unión de HIF a sus genes diana, conlleva una alteración de la regulación de los mismos. Como hemos planteado, la presencia de SNPs en los elementos de respuesta a hipoxia podría tener importantes implicaciones en respuestas alteradas en enfermedades que cursan con hipoxia.

## ***DISCUSIÓN***



Para tener una visión global de la respuesta celular a hipoxia es fundamental conocer los cambios transcripcionales mediados por HIF, puesto que es el principal efector de la misma. Con este objetivo, en el presente trabajo, hemos desarrollado una estrategia bioinformática, de alta especificidad y sensibilidad, basada en datos experimentales de perfiles de expresión génica y en la localización *in silico* de los elementos de unión de HIF en todo el genoma, obteniendo así un listado de nuevas dianas de HIF. Por otro lado, tener las coordenadas genómicas relativas a los potenciales HBSs nos permite identificar SNPs que pueden alterar la respuesta a hipoxia.

En los últimos años se han desarrollado numerosas técnicas experimentales de alto rendimiento (high-throughput) con las que se trata de identificar todo el conjunto de genes que se modulan por un factor de transcripción determinado o ante un estímulo. Un ejemplo de ello son los perfiles de expresión génica, éstos ofrecen una gran cantidad de información sobre las modulaciones transcripcionales asociadas un tratamiento determinado; sin embargo, una importante limitación de esta aproximación experimental es, que no permite determinar si estos cambios están mediados por un factor de transcripción en concreto, ni si se trata de regulaciones directas o indirectas mediadas por este factor. Por otro lado, estas técnicas son muy ruidosas, tal y como queda reflejado en este trabajo, y muy dependientes de las condiciones experimentales utilizadas (tipo celular, tipo de estímulo, plataforma, etc.). Por tanto, sin un análisis global (meta-análisis) de los datos como el que hemos realizado, no es posible extraer un conjunto robusto de genes que se module por hipoxia, sin incluir gran cantidad de falsos positivos debido al ruido intrínseco asociado a los propios perfiles de expresión, o por el contrario perder gran cantidad de dianas reales consecuencia de la intersección de los datos. En este trabajo demostramos que el meta-análisis permite extraer un conjunto fidedigno de genes modulados por hipoxia salvando las dos limitaciones anteriores. Esto queda reflejado cuando comparamos los análisis individuales de los perfiles de expresión y el meta-análisis, utilizando un conjunto de genes diana conocidos de HIF (fig 15). Una conclusión llamativa es que, a pesar de que el meta-análisis tiene una mayor sensibilidad, existen varios genes que aparecen como modulados por hipoxia en la intersección de los datos de los análisis individuales de los perfiles de expresión, como son NDRG1, HIST2HBE, PDK1, BHLHB2, C3ORF28, ZNF395 y CCNG2 (tabla 7) que no aparecen en el meta-análisis. A priori parece que estos genes deberían estar presentes en ambos tipos de análisis, sin embargo, los diferentes criterios a la hora de considerar a un gen concreto como modulado o no varían en cada tipo de análisis. Así, en los análisis de experimentos individuales consideramos como modulados a aquellos genes que tengan al menos una de sus sondas asociadas, modulada por

## Discusión

hipoxia ( $\log Hx/Nx > 1.96SD$ ; ( $\log Hx/Nx > 2.6 SD$ ), mientras que en el meta-análisis no solo tenemos en cuenta esta modulación ( $\log Hx/Nx$ ) sino también la intensidad de esta modulación (p-valor ajustado), esto nos permite descartar variaciones no significativas.

Otro de los abordajes experimentales de alto rendimiento, encaminado a identificar genes diana de un factor de transcripción, son los experimentos de CHIP-CHIP. Estos permiten la identificación a gran escala de todos los sitios de unión de un factor de transcripción en la totalidad del genoma. Estos abordajes son muy informativos, pero al igual que pasa con los perfiles de expresión génica, pueden ser altamente ruidosos. Claro reflejo de ello es el poco solapamiento existente la intersección de los datos de dos trabajos independientes de CHIP-CHIP para HIF1 $\alpha$  (fig 16A) (Mole, Blancher et al. 2009; Xia, Lemieux et al. 2009), esto viene dado no solo por las diferentes líneas celulares y plataformas de hibridación utilizadas, sino también por el ruido intrínseco a la propia técnica.

La identificación de elementos de respuesta de un factor de transcripción utilizando técnicas computacionales es un abordaje cada vez más utilizado, ya que permite elaborar finas predicciones salvando los costes de los abordajes experimentales de alto rendimiento. En nuestro trabajo, la identificación de HBSs se basa en dos premisas con las que tratamos de minimizar la cantidad de HBSs no funcionales, algo necesario puesto que la secuencia del núcleo central del HRE (RCGTG) por su pequeño tamaño, puede encontrarse por azar en cada uno de los genes del genoma al menos una vez. Así, restringimos nuestra búsqueda a secuencias del genoma no codificantes, puesto que es en ellas donde se localizan la mayor parte de los elementos reguladores de la transcripción (Visel, Rubin et al. 2009), y a regiones conservadas evolutivamente (PhasCons) (Siepel, Bejerano et al. 2005), ya que el criterio de conservación supone que las regiones reguladoras tienden a conservarse evolutivamente en mayor medida frente al resto. Así en este trabajo definimos CNSs (*Conserved Non-coding Sequences*) a las que restringimos la búsqueda de secuencias RCGTG, criterio que parece ser acertado, puesto que cerca del 80% de los HBSs identificados experimentalmente (Xia, Lemieux et al. 2009) se localizan en CNSs.

Tras la identificación de los sitios RCGTG, la herramienta bioinformática les asigna una puntuación que es indicativa de su similitud con secuencias HRE funcionales. Para ello comparamos estas secuencias con una matriz de frecuencias (PWM) elaborada a partir de genes diana de HIF bien caracterizados (anexo II). La PWM no solo contiene la secuencia del núcleo central RCGTG, sino que consideramos también importantes las posiciones adyacentes. El alineamiento de las 46 secuencias del HRE de genes diana conocidos de HIF reveló que

ciertas bases en torno al RCGTG se presentan con mayor probabilidad a la esperada por azar. Así, consideramos que estas bases son necesarias para la actividad transcripción de HIF, bien porque se pueda unir algún coactivador transcripcional o bien porque esas bases sean necesarias para la unión de HIF $\alpha$  o HIF $\beta$  o porque confieran una estructura favorable para la unión del complejo transcripcional HIF. El valor predictivo de la puntuación asignada en base a esta PWM está avalado por el hecho de que, los HBSs identificados experimentalmente en ensayos de ChIP-ChIP (Mole, Blancher et al. 2009; Xia, Lemieux et al. 2009), tienen puntuaciones claramente superiores a las de las regiones control, nuestra PWM les asigna los valores más elevados (fig 16 B y C).

La combinación de los datos experimentales de perfiles de expresión génica junto con la identificación *in silico* de los HBSs, nos permite integrar la información de las aproximaciones experimentales y bioinformáticas en una estrategia con la que obtenemos un listado fiel de genes diana de HIF. A lo largo del desarrollo de toda la estrategia hemos seleccionado unos puntos de corte muy estrictos, tanto en el meta-análisis (tomamos como modulados los genes que tienen un p-valor ajustado < 0.01) como en la estrategia final (Pt/Pb=6.5). Así, identificamos dianas de HIF con una sensibilidad del 78%, manteniendo una elevada especificidad (97.77%). Como consecuencia de la astringencia de estos criterios perdemos, en el listado final, genes diana de HIF bien caracterizados como EPO. Puede parecer paradójico que una estrategia que parece tener alta sensibilidad y especificidad no sea capaz de detectar la primera diana de HIF descrita. Sin embargo, este hecho es debido a que EPO se expresa de forma tejida específica (tan solo en algunos tipos celulares hepáticos y renales), y por tanto su representación es muy minoritaria en los perfiles de expresión génica. Con el fin de minimizar la pérdida de genes como EPO en el listado de potenciales dianas de HIF, equilibramos la contribución de la modulación de cada gen por hipoxia teniendo en cuenta la consistencia de su modulación (p-valor) en la puntuación final Pt/Pb. Así EPO, a pesar de no estar considerado en nuestro listado como diana de HIF, está posicionado muy próximo al punto de corte (Pt/Pb= 5.17), de hecho aparece en la posición 360 de 11672 genes analizados. Los estrictos criterios aplicados a la estrategia nos permiten reducir tremendamente la cantidad de falsos positivos y, consideramos la pérdida de genes como EPO, un precio justo por una estrategia más fiable. De hecho, la validación experimental mediante ensayos de ChIP refleja que, con los parámetros adoptados, la estrategia es capaz de clasificar correctamente genes en dianas y no dianas de HIF con una eficiencia del 80%. Por tanto, este abordaje es tan fiable o más que ensayos experimentales en la detección de genes diana de HIF, suponiendo además una importante reducción de la inversión de tiempo y dinero.

Tras la aplicación de esta estrategia obtuvimos un listado de 216 potenciales dianas de HIF, número similar al conjunto de genes diana identificados en los abordajes experimentales de CHIP-CHIP para HIF1 $\alpha$  (Mole, Blancher et al. 2009; Xia, Lemieux et al. 2009). Sin embargo, pensamos que nuestra valoración subestima el número de dianas reales de HIF, puesto que, para el desarrollo de la estrategia, nos ceñimos solo a genes que tuvieran asociada información experimental (perfiles de expresión de GEO). El estudio se realizó con 11.000 genes de los aproximadamente 25.000 del genoma, lo que supone tan solo el 50%. Por lo tanto, estimamos que el número total de dianas de HIF podría ser en torno a 500 genes, lo que supone un 1% de la totalidad del genoma. Por otro lado, vemos que la mayoría de los genes diana de HIF bien caracterizados se posicionan en los primeros puestos del listado (anexo IV), lo que supone un grado más de confianza en la fiabilidad de la estrategia desarrollada.

Una característica importante del abordaje seguido es que no hacemos distinciones en cuanto a la isoforma de HIF $\alpha$  implicada en la regulación transcripcional, de hecho los experimentos de GEO son realizados en condiciones de hipoxia o con miméticos de hipoxia, que estabilizan ambas isoformas. Solo existe un experimento en el que se sobreexpresan selectivamente una de las isoformas de HIF $\alpha$  (GSE2020). Además, en la identificación de HBSs solo tenemos en cuenta la secuencia de unión de HIF (RCGTG) a la que, tal y como se ha validado experimentalmente (Mole, Blancher et al. 2009), se unen por igual las dos isoformas de HIF $\alpha$ . Como hemos mencionado a lo largo del texto, numerosos trabajos proponen firmemente que la especificidad de sustrato depende de la interacción de HIF con cofactores transcripcionales, por tanto, el listado de dianas englobaría el conjunto de genes regulados por cada isoforma de forma independiente y los genes comunes a ambas.

El análisis de ontología del listado de potenciales dianas de HIF, puso de manifiesto funciones biológicas ya asociadas a este factor como es el metabolismo de la glucosa, y además reveló una nueva función, el metabolismo el glucógeno, hasta ahora no asociada a HIF. El acúmulo de glucógeno en hipoxia es algo que se ha descrito en la bibliografía de forma muy puntual y descriptiva (Vigoda, Mamedova et al. 2003) (Mamedova, Shneyvays et al. 2003), sin profundizar en los mecanismo moleculares subyacentes. En este trabajo describimos cuatro genes diana de HIF, que controlan en varios puntos clave el proceso de síntesis de glucógeno, como son: GYS1, sintasa que cataliza la síntesis de glucógeno a partir de UDP-glucosa; UGP2, que sintetiza UDP-glucosa, sustrato de GYS1; GBE1 que permite la ramificación del polímero de glucógeno, proceso esencial para su función biológicas; y el enzima PPP1R3C que es la subunidad reguladora de PP1 y que activa la GYS. Así pues, en hipoxia, HIF está mediando la



inducción de todos estos enzimas, potenciando el acúmulo de glucógeno a través de la síntesis del mismo.

En hipoxia, la activación de HIF promueve una reducción importante de la fosforilación oxidativa mediante la inducción de PDK1. La respuesta mediada por HIF, trata de compensar este déficit energético que tiene lugar en hipoxia, potenciando el flujo a través de las vías anaerobias. Para ello, induce prácticamente todos los enzimas implicados en glucólisis y fermentación. Parece por tanto, algo contradictorio, que la célula en esta situación invierta glucosa en un proceso de síntesis de glucógeno en lugar de degradarlo para obtener ATP. Uno de los mecanismos que podría explicar este hecho es que, en los primeros minutos de hipoxia, se produjese un rápido consumo de todas las reservas de glucógeno de la célula y que la inducción de la síntesis de glucógeno mediada por HIF estuviese orientada a compensar este desgaste, recuperando así los niveles basales de reserva de glucógeno (Mamedova, Shneyvays et al. 2003). Además, el acúmulo de glucógeno podría estar integrado en el proceso de acondicionamiento hipóxico (Aragones, Fraisl et al. 2009), descrito en tejidos como corazón, cerebro, riñón, etc (Bernhardt, Campean et al. 2006) (Cai, Zhong et al. 2008) (Helton, Cui et al. 2005) (Ran, Xu et al. 2005). Durante este proceso, todos los cambios metabólicos que tienen lugar van encaminados a preparar la célula para una situación de hipoxia posterior más aguda o prolongada, permitiendo la mejor adaptación celular a estas situaciones de estrés posterior así como la supervivencia celular (Pescador, Villar et al. 2010). Por otro lado, este acumulo de glucógeno podría tener una explicación más sencilla, podría tratarse, simplemente, de un mecanismo de compensación metabólica; en hipoxia, HIF fuerza la entrada en la célula de gran cantidad de glucosa induciendo la expresión del transportador GLUT1, si toda esta glucosa que entra en la célula no puede ser absorbida por la vía glucolítica se acumularía en el citoplasma, llevando a un estado de desequilibrio osmótico, así, la entrada de esta glucosa libre en la vía de la gluconeogénesis equilibraría esta situación.

Otra de las funciones biológicas sobrerrepresentadas en el análisis de ontología es la regulación de la transcripción génica, HIF induce genes implicados en regulación transcripcional como: CITED2, IRF1, MAFF, HOXA4, etc. Entre ellos, destacan varios factores de transcripción que participan en represión transcripcional, como es el caso de RCOR2 cuya regulación dependiente de HIF ha sido verificada en nuestro laboratorio (Ortiz-Barahona, Villar et al. 2010). Muchos genes se reprimen en hipoxia, sin embargo, el resultado del análisis del listado total de genes modulados por hipoxia muestran una ausencia de correlación entre la presencia de HBS y la represión transcripcional. Este resultado ha sido corroborado en otros trabajos experimentales (Mole, Blancher et al. 2009). Por tanto, pensamos que esta represión

## Discusión

transcripcional podría estar dirigida por represores transcripcionales inducidos vía HIF en hipoxia. Otro mecanismo que podría explicar esta represión es la inducción de miRNAs durante la hipoxia, se ha visto que HIF es capaz de inducir miRNAs que promueven la bajada en la expresión de determinados genes. Esta cuestión ha sido poco tratada en el campo, por tanto, consideramos que sería muy interesante profundizar en los mecanismos subyacentes a esta represión transcripcional medida por hipoxia.

Finalmente destacar que, a pesar de que la angiogénesis es una de las funciones biológicas clásicamente asociadas con la hipoxia, no aparece representada en el análisis de ontología realizado. Esto es debido a que el número de genes diana que representan esta función biológica es despreciable frente a los otros grupos funcionales mostrados (fig 21).

Por otro lado, la estrategia desarrollada nos permite obtener la información relativa a las coordenadas genómicas globales de HBSs, esto no solo nos sirve para predecir dianas directas de HIF, sino que también nos permite identificar polimorfismos en ellas que puedan afectar a la regulación transcripcional mediada por HIF. El análisis de SNPs demuestra que existen 12 variaciones polimórficas que se localizan en el núcleo central (RCGTG) del HBS de genes clasificados como diana de HIF. El estudio funcional del SNP de MIF (factor inhibidor de macrófagos) reveló la importancia de la integridad de este núcleo RCGTG en la modulación transcripcional mediada por HIF. MIF, es una citoquina que juega un papel principal en inmunidad innata y adquirida (Cooke, Armstrong et al. 2009), en inmunoregulación e inflamación (Takahashi, Koga et al. 2009); también se ha asociado con ciertas patologías como alteraciones del sistema inmune, entre las que destacan artritis reumatoide (Lue, Kleemann et al. 2002) y sepsis (Parrish, Gallowitsch-Puerta et al. 2008), y, en los últimos años, ha sido relacionado con carcinogénesis y metástasis (He, Chen et al. 2009). Debido a que HIF modula gran cantidad de genes implicados en progresión tumoral, este gen supuso un modelo atractivo de estudio. Nos planteamos si, quizá, la presencia de SNPs en sitios HRE podría tener alguna implicación en patologías que cursan con hipoxia, consecuencia de una respuesta alterada a este estímulo. Los ensayos llevados a cabo con construcciones que mimetizan el SNP en el sitio HRE de MIF, pusieron de manifiesto una desregulación de la transcripción mediada por HIF en hipoxia. Es difícil tratar de predecir la implicación de la presencia del polimorfismo en el HRE de MIF puesto que, probablemente, sería necesario que éste se diese en homocigosis para ver algún efecto claro en la regulación de MIF, de hecho, la eliminación de el HRE de un gen regulado por hipoxia puede tener consecuencias tan dramáticas como se demostró para el caso de VEGF (Oosthuysse, Moons et al. 2001) (Vinores, Xiao et al. 2006). En cualquier caso, esta es la primera evidencia de la existencia de variaciones en la respuesta a

hipoxia entre individuos asociada a la presencia de SNPs. El conocimiento de la existencia de estas variaciones entre individuos, puede ser un importante campo de estudio a la hora de generar terapias personalizadas en enfermedades que cursan con hipoxia. Por último, estos resultados evidencian la importancia de las mutaciones en elementos reguladores, no solo las mutaciones en secuencias codificantes son las responsables de las variaciones entre individuos.

En conjunto, todos los datos presentados en este trabajo muestran que, la combinación de datos experimentales de perfiles de expresión génica junto con la identificación *in silico* de sitios de unión de HIF en todo el genoma es una estrategia muy fiable para identificar nuevas dianas directas de HIF. De hecho, los dos trabajos de ChIP-ChIP que utilizamos como comparativa a lo largo del trabajo muestran muy poco solapamiento entre ellos, solo 37 de ellos son solapantes (tabla 8), lo que refleja la importancia de las variaciones experimentales en este tipo de ensayos, y esta es una limitación que queda subsanada en nuestra estrategia. Con ella, no solo conseguimos identificar 152 genes diana de HIF no descritos, sino que revelamos una nueva función que parece tener relevancia en la adaptación metabólica celular a la hipoxia. Finalmente, esta estrategia nos permite identificar polimorfismos que alteran la respuesta a hipoxia y cuya implicación en enfermedades podría abrir nuevos frentes en el desarrollo de tratamientos farmacológicos dirigidos.



## ***CONCLUSIONES***



1. Hemos desarrollado una estrategia bioinformática que, combinando datos de expresión génica junto con la identificación de sitios de unión de HIF, nos permite identificar genes diana de HIF con una elevada especificidad y sensibilidad.
2. La aplicación de esta estrategia nos permitió obtener un listado fiel de potenciales genes diana de HIF. Las validaciones experimentales mediante ensayos de ChIP, revelaron que nuestra estrategia clasifica genes como diana o no diana de HIF correctamente en un 80% de los casos.
3. El análisis de nuestro listado puso de manifiesto que HIF, de forma global, media la inducción de genes en respuesta a hipoxia y no la represión. No existe asociación entre represión génica por hipoxia y presencia de sitios de unión de HIF (HBSs), esta conclusión sin embargo, no descarta la posibilidad de casos concretos de represión transcripcional mediada por HIF.
4. El análisis de ontología del listado de potenciales dianas de HIF reveló que, además de las funciones biológicas clásicamente asociadas a HIF en respuesta a hipoxia, aparecía una nueva función previamente no descrita, la síntesis de glucógeno.
5. Cuatro nuevas dianas de HIF están implicadas en la síntesis de glucógeno: GYS1, UGP2, PPP1R3C y GBE1. La validación funcional de estos genes evidenció su regulación transcripcional mediada por HIF.
6. Una extensión de la estrategia bioinformática permitió identificar 12 SNPs que se localizaban dentro de las coordenadas genómicas de los HBSs de genes clasificados como dianas de HIF.
7. El estudio funcional del SNP localizado en el HRE del locus de MIF reflejó la importancia de la integridad del núcleo central del HRE (RCGTG) en la regulación transcripcional mediada por HIF. Ésta es la primera evidencia de la asociación entre variaciones genéticas y alteración de la regulación transcripcional.





## ***BIBLIOGRAFÍA***



- Appelhoff, R. J., Y. M. Tian, et al. (2004). "Differential function of the prolyl hydroxylases PHD1, PHD2, and PHD3 in the regulation of hypoxia-inducible factor." J Biol Chem **279**(37): 38458-38465.
- Aprelikova, O., M. Wood, et al. (2006). "Role of ETS transcription factors in the hypoxia-inducible factor-2 target gene selection." Cancer Res **66**(11): 5641-5647.
- Aragones, J., P. Fraisl, et al. (2009). "Oxygen sensors at the crossroad of metabolism." Cell Metab **9**(1): 11-22.
- Aragones, J., M. Schneider, et al. (2008). "Deficiency or inhibition of oxygen sensor Phd1 induces hypoxia tolerance by reprogramming basal metabolism." Nat Genet **40**(2): 170-180.
- Barth, S., J. Nesper, et al. (2007). "The peptidyl prolyl cis/trans isomerase FKBP38 determines hypoxia-inducible transcription factor prolyl-4-hydroxylase PHD2 protein stability." Mol Cell Biol **27**(10): 3758-3768.
- Baugh, J. A., M. Gantier, et al. (2006). "Dual regulation of macrophage migration inhibitory factor (MIF) expression in hypoxia by CREB and HIF-1." Biochem Biophys Res Commun **347**(4): 895-903.
- Bernhardt, W. M., V. Campean, et al. (2006). "Preconditional activation of hypoxia-inducible factors ameliorates ischemic acute renal failure." J Am Soc Nephrol **17**(7): 1970-1978.
- Berra, E., E. Benizri, et al. (2003). "HIF prolyl-hydroxylase 2 is the key oxygen sensor setting low steady-state levels of HIF-1alpha in normoxia." EMBO J **22**(16): 4082-4090.
- Berra, E., D. E. Richard, et al. (2001). "HIF-1-dependent transcriptional activity is required for oxygen-mediated HIF-1alpha degradation." FEBS Lett **491**(1-2): 85-90.
- Berta, M. A., N. Mazure, et al. (2007). "SUMOylation of hypoxia-inducible factor-1alpha reduces its transcriptional activity." Biochem Biophys Res Commun **360**(3): 646-652.
- Bishop, T., D. Gallagher, et al. (2008). "Abnormal sympathoadrenal development and systemic hypotension in PHD3<sup>-/-</sup> mice." Mol Cell Biol **28**(10): 3386-3400.
- Blanchette, M. and S. Sinha (2001). "Separating real motifs from their artifacts." Bioinformatics **17 Suppl 1**: S30-38.
- Bracken, C. P., M. L. Whitelaw, et al. (2005). "Activity of hypoxia-inducible factor 2alpha is regulated by association with the NF-kappaB essential modulator." J Biol Chem **280**(14): 14240-14251.
- Bruick, R. K. and S. L. McKnight (2001). "A conserved family of prolyl-4-hydroxylases that modify HIF." Science **294**(5545): 1337-1340.
- Cai, Z., H. Zhong, et al. (2008). "Complete loss of ischaemic preconditioning-induced cardioprotection in mice with partial deficiency of HIF-1 alpha." Cardiovasc Res **77**(3): 463-470.
- Carrero, P., K. Okamoto, et al. (2000). "Redox-regulated recruitment of the transcriptional coactivators CREB-binding protein and SRC-1 to hypoxia-inducible factor 1alpha." Mol Cell Biol **20**(1): 402-415.
- Compernelle, V., K. Brusselmans, et al. (2002). "Loss of HIF-2alpha and inhibition of VEGF impair fetal lung maturation, whereas treatment with VEGF prevents fatal respiratory distress in premature mice." Nat Med **8**(7): 702-710.
- Cooke, G., M. E. Armstrong, et al. (2009). "Macrophage migration inhibitory factor (MIF), enzymatic activity and the inflammatory response." Biofactors **35**(2): 165-168.
- Covello, K. L., J. Kehler, et al. (2006). "HIF-2alpha regulates Oct-4: effects of hypoxia on stem cell function, embryonic development, and tumor growth." Genes Dev **20**(5): 557-570.
- Cramer, T., Y. Yamanishi, et al. (2003). "HIF-1alpha is essential for myeloid cell-mediated inflammation." Cell **112**(5): 645-657.
- Chen, K. F., Y. Y. Lai, et al. (2005). "Transcriptional repression of human cad gene by hypoxia inducible factor-1alpha." Nucleic Acids Res **33**(16): 5190-5198.
- Damert, A., E. Ikeda, et al. (1997). "Activator-protein-1 binding potentiates the hypoxia-inducible factor-1-mediated hypoxia-induced transcriptional activation of vascular-

- endothelial growth factor expression in C6 glioma cells." Biochem J **327 ( Pt 2)**: 419-423.
- Davidson, E. H., J. P. Rast, et al. (2002). "A genomic regulatory network for development." Science **295(5560)**: 1669-1678.
- Dayan, F., D. Roux, et al. (2006). "The oxygen sensor factor-inhibiting hypoxia-inducible factor-1 controls expression of distinct genes through the bifunctional transcriptional character of hypoxia-inducible factor-1alpha." Cancer Res **66(7)**: 3688-3698.
- del Peso, L., M. C. Castellanos, et al. (2003). "The von Hippel Lindau/hypoxia-inducible factor (HIF) pathway regulates the transcription of the HIF-proline hydroxylase genes in response to low oxygen." J Biol Chem **278(49)**: 48690-48695.
- Ebert, B. L. and H. F. Bunn (1998). "Regulation of transcription by hypoxia requires a multiprotein complex that includes hypoxia-inducible factor 1, an adjacent transcription factor, and p300/CREB binding protein." Mol Cell Biol **18(7)**: 4089-4096.
- Ehrismann, D., E. Flashman, et al. (2007). "Studies on the activity of the hypoxia-inducible-factor hydroxylases using an oxygen consumption assay." Biochem J **401(1)**: 227-234.
- Elgar, G. and T. Vavouri (2008). "Tuning in to the signals: noncoding sequence conservation in vertebrate genomes." Trends Genet **24(7)**: 344-352.
- Elvert, G., A. Kappel, et al. (2003). "Cooperative interaction of hypoxia-inducible factor-2alpha (HIF-2alpha ) and Ets-1 in the transcriptional activation of vascular endothelial growth factor receptor-2 (Flk-1)." J Biol Chem **278(9)**: 7520-7530.
- Epstein, A. C., J. M. Gleadle, et al. (2001). "C. elegans EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation." Cell **107(1)**: 43-54.
- Feldser, D., F. Agani, et al. (1999). "Reciprocal positive regulation of hypoxia-inducible factor 1alpha and insulin-like growth factor 2." Cancer Res **59(16)**: 3915-3918.
- Firth, J. D., B. L. Ebert, et al. (1995). "Hypoxic regulation of lactate dehydrogenase A. Interaction between hypoxia-inducible factor 1 and cAMP response elements." J Biol Chem **270(36)**: 21021-21027.
- Freedman, S. J., Z. Y. Sun, et al. (2003). "Structural basis for negative regulation of hypoxia-inducible factor-1alpha by CITED2." Nat Struct Biol **10(7)**: 504-512.
- Freeman, R. S., D. M. Hasbani, et al. (2003). "SM-20, EGL-9, and the EGLN family of hypoxia-inducible factor prolyl hydroxylases." Mol Cells **16(1)**: 1-12.
- Gordan, J. D., J. A. Bertout, et al. (2007). "HIF-2alpha promotes hypoxic cell proliferation by enhancing c-myc transcriptional activity." Cancer Cell **11(4)**: 335-347.
- Gordan, J. D., C. B. Thompson, et al. (2007). "HIF and c-Myc: sibling rivals for control of cancer cell metabolism and proliferation." Cancer Cell **12(2)**: 108-113.
- Gray, M. J., J. Zhang, et al. (2005). "HIF-1alpha, STAT3, CBP/p300 and Ref-1/APE are components of a transcriptional complex that regulates Src-dependent hypoxia-induced expression of VEGF in pancreatic and prostate carcinomas." Oncogene **24(19)**: 3110-3120.
- Gruber, M., C. J. Hu, et al. (2007). "Acute postnatal ablation of Hif-2alpha results in anemia." Proc Natl Acad Sci U S A **104(7)**: 2301-2306.
- Gustafsson, M. V., X. Zheng, et al. (2005). "Hypoxia requires notch signaling to maintain the undifferentiated cell state." Dev Cell **9(5)**: 617-628.
- He, X. X., K. Chen, et al. (2009). "Macrophage migration inhibitory factor promotes colorectal cancer." Mol Med **15(1-2)**: 1-10.
- Helton, R., J. Cui, et al. (2005). "Brain-specific knock-out of hypoxia-inducible factor-1alpha reduces rather than increases hypoxic-ischemic damage." J Neurosci **25(16)**: 4099-4107.
- Hirsila, M., P. Koivunen, et al. (2003). "Characterization of the human prolyl 4-hydroxylases that modify the hypoxia-inducible factor." J Biol Chem **278(33)**: 30772-30780.
- Hoffman, E. C., H. Reyes, et al. (1991). "Cloning of a factor required for activity of the Ah (dioxin) receptor." Science **252(5008)**: 954-958.

- Hsieh, M. M., N. S. Linde, et al. (2007). "HIF prolyl hydroxylase inhibition results in endogenous erythropoietin induction, erythrocytosis, and modest fetal hemoglobin expression in rhesus macaques." Blood **110**(6): 2140-2147.
- Hu, C. J., A. Sataur, et al. (2007). "The N-terminal transactivation domain confers target gene specificity of hypoxia-inducible factors HIF-1alpha and HIF-2alpha." Mol Biol Cell **18**(11): 4528-4542.
- Hu, C. J., L. Y. Wang, et al. (2003). "Differential roles of hypoxia-inducible factor 1alpha (HIF-1alpha) and HIF-2alpha in hypoxic gene regulation." Mol Cell Biol **23**(24): 9361-9374.
- Huang da, W., B. T. Sherman, et al. (2009). "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources." Nat Protoc **4**(1): 44-57.
- Huang, L. E., Z. Arany, et al. (1996). "Activation of hypoxia-inducible transcription factor depends primarily upon redox-sensitive stabilization of its alpha subunit." J Biol Chem **271**(50): 32253-32259.
- Huang, L. E., J. Gu, et al. (1998). "Regulation of hypoxia-inducible factor 1alpha is mediated by an O<sub>2</sub>-dependent degradation domain via the ubiquitin-proteasome pathway." Proc Natl Acad Sci U S A **95**(14): 7987-7992.
- Huang, Y., R. P. Hickey, et al. (2004). "Cardiac myocyte-specific HIF-1alpha deletion alters vascularization, energy availability, calcium flux, and contractility in the normoxic heart." FASEB J **18**(10): 1138-1140.
- Ivan, M., K. Kondo, et al. (2001). "HIFalpha targeted for VHL-mediated destruction by proline hydroxylation: implications for O<sub>2</sub> sensing." Science **292**(5516): 464-468.
- Iyer, N. V., L. E. Kotch, et al. (1998). "Cellular and developmental control of O<sub>2</sub> homeostasis by hypoxia-inducible factor 1 alpha." Genes Dev **12**(2): 149-162.
- Iyer, N. V., S. W. Leung, et al. (1998). "The human hypoxia-inducible factor 1alpha gene: HIF1A structure and evolutionary conservation." Genomics **52**(2): 159-165.
- Jaakkola, P., D. R. Mole, et al. (2001). "Targeting of HIF-alpha to the von Hippel-Lindau ubiquitylation complex by O<sub>2</sub>-regulated prolyl hydroxylation." Science **292**(5516): 468-472.
- Jeong, J. W., M. K. Bae, et al. (2002). "Regulation and destabilization of HIF-1alpha by ARD1-mediated acetylation." Cell **111**(5): 709-720.
- Jiang, B. H., J. Z. Zheng, et al. (1997). "Transactivation and inhibitory domains of hypoxia-inducible factor 1alpha. Modulation of transcriptional activity by oxygen tension." J Biol Chem **272**(31): 19253-19260.
- Kaelin, W. G. (2007). "Von Hippel-Lindau disease." Annu Rev Pathol **2**: 145-173.
- Koivunen, P., P. Tiainen, et al. (2007). "An endoplasmic reticulum transmembrane prolyl 4-hydroxylase is induced by hypoxia and acts on hypoxia-inducible factor alpha." J Biol Chem **282**(42): 30544-30552.
- Kozak, K. R., B. Abbott, et al. (1997). "ARNT-deficient mice and placental differentiation." Dev Biol **191**(2): 297-305.
- Laderoute, K. R. (2005). "The interaction between HIF-1 and AP-1 transcription factors in response to low oxygen." Semin Cell Dev Biol **16**(4-5): 502-513.
- Landazuri, M. O., A. Vara-Vega, et al. (2006). "Analysis of HIF-prolyl hydroxylases binding to substrates." Biochem Biophys Res Commun **351**(2): 313-320.
- Lando, D., D. J. Peet, et al. (2002). "FIH-1 is an asparaginyl hydroxylase enzyme that regulates the transcriptional activity of hypoxia-inducible factor." Genes Dev **16**(12): 1466-1471.
- Lando, D., D. J. Peet, et al. (2002). "Asparagine hydroxylation of the HIF transactivation domain a hypoxic switch." Science **295**(5556): 858-861.
- Lieb, M. E., K. Menzies, et al. (2002). "Mammalian EGLN genes have distinct patterns of mRNA expression and regulation." Biochem Cell Biol **80**(4): 421-426.
- Lisy, K. and D. J. Peet (2008). "Turn me on: regulating HIF transcriptional activity." Cell Death Differ **15**(4): 642-649.

- Little, P. F. (2005). "Structure and function of the human genome." Genome Res **15**(12): 1759-1766.
- Liu, Y. V., J. H. Baek, et al. (2007). "RACK1 competes with HSP90 for binding to HIF-1alpha and is required for O(2)-independent and HSP90 inhibitor-induced degradation of HIF-1alpha." Mol Cell **25**(2): 207-217.
- Lopez-Barneo, J., P. Ortega-Saenz, et al. (2009). "Oxygen sensing in the carotid body." Ann N Y Acad Sci **1177**: 119-131.
- Lue, H., R. Kleemann, et al. (2002). "Macrophage migration inhibitory factor (MIF): mechanisms of action and role in disease." Microbes Infect **4**(4): 449-460.
- Luo, W., J. Zhong, et al. (2010). "Hsp70 and CHIP selectively mediate ubiquitination and degradation of hypoxia-inducible factor (HIF)-1alpha but Not HIF-2alpha." J Biol Chem **285**(6): 3651-3663.
- Makino, Y., R. Cao, et al. (2001). "Inhibitory PAS domain protein is a negative regulator of hypoxia-inducible gene expression." Nature **414**(6863): 550-554.
- Mamedova, L. K., V. Shneyvays, et al. (2003). "Mechanism of glycogen supercompensation in rat skeletal muscle cultures." Mol Cell Biochem **250**(1-2): 11-19.
- Mason, S. D., R. A. Howlett, et al. (2004). "Loss of skeletal muscle HIF-1alpha results in altered exercise endurance." PLoS Biol **2**(10): e288.
- Maxwell, P. H., M. S. Wiesener, et al. (1999). "The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis." Nature **399**(6733): 271-275.
- Metzen, E., U. Berchner-Pfannschmidt, et al. (2003). "Intracellular localisation of human HIF-1 alpha hydroxylases: implications for oxygen sensing." J Cell Sci **116**(Pt 7): 1319-1326.
- Metzen, E., D. P. Stiehl, et al. (2005). "Regulation of the prolyl hydroxylase domain protein 2 (phd2/egln-1) gene: identification of a functional hypoxia-responsive element." Biochem J **387**(Pt 3): 711-717.
- Mole, D. R., C. Blancher, et al. (2009). "Genome-wide association of hypoxia-inducible factor (HIF)-1alpha and HIF-2alpha DNA binding with expression profiling of hypoxia-inducible transcripts." J Biol Chem **284**(25): 16767-16775.
- Mylonis, I., G. Chachami, et al. (2006). "Identification of MAPK phosphorylation sites and their role in the localization and activity of hypoxia-inducible factor-1alpha." J Biol Chem **281**(44): 33095-33106.
- Nakayama, K., I. J. Frew, et al. (2004). "Siah2 regulates stability of prolyl-hydroxylases, controls HIF1alpha abundance, and modulates physiological responses to hypoxia." Cell **117**(7): 941-952.
- O'Rourke, J. F., Y. M. Tian, et al. (1999). "Oxygen-regulated and transactivating domains in endothelial PAS protein 1: comparison with hypoxia-inducible factor-1alpha." J Biol Chem **274**(4): 2060-2071.
- Oosthuysen, B., L. Moons, et al. (2001). "Deletion of the hypoxia-response element in the vascular endothelial growth factor promoter causes motor neuron degeneration." Nat Genet **28**(2): 131-138.
- Ortiz-Barahona, A., D. Villar, et al. (2010). "Genome-wide identification of hypoxia-inducible factor binding sites and target genes by a probabilistic model integrating transcription-profiling data and in silico binding site prediction." Nucleic Acids Res **38**(7): 2332-2345.
- Parrish, W. R., M. Gallowitsch-Puerta, et al. (2008). "Experimental therapeutic strategies for severe sepsis: mediators and mechanisms." Ann N Y Acad Sci **1144**: 210-236.
- Peng, J., L. Zhang, et al. (2000). "The transcription factor EPAS-1/hypoxia-inducible factor 2alpha plays an important role in vascular remodeling." Proc Natl Acad Sci U S A **97**(15): 8386-8391.
- Percy, M. J., Q. Zhao, et al. (2006). "A family with erythrocytosis establishes a role for prolyl hydroxylase domain protein 2 in oxygen homeostasis." Proc Natl Acad Sci U S A **103**(3): 654-659.

- Perkins, N. D. and T. D. Gilmore (2006). "Good cop, bad cop: the different faces of NF-kappaB." Cell Death Differ **13**(5): 759-772.
- Pescador, N., Y. Cuevas, et al. (2005). "Identification of a functional hypoxia-responsive element that regulates the expression of the egl nine homologue 3 (egln3/phd3) gene." Biochem J **390**(Pt 1): 189-197.
- Pescador, N., D. Villar, et al. (2010). "Hypoxia promotes glycogen accumulation through hypoxia inducible factor (HIF)-mediated induction of glycogen synthase 1." PLoS One **5**(3): e9644.
- Peyssonaux, C., V. Datta, et al. (2005). "HIF-1alpha expression regulates the bactericidal capacity of phagocytes." J Clin Invest **115**(7): 1806-1815.
- Ran, R., H. Xu, et al. (2005). "Hypoxia preconditioning in the brain." Dev Neurosci **27**(2-4): 87-92.
- Ratan, R. R., A. Siddiq, et al. (2007). "Harnessing hypoxic adaptation to prevent, treat, and repair stroke." J Mol Med **85**(12): 1331-1338.
- Richard, D. E., E. Berra, et al. (2000). "Nonhypoxic pathway mediates the induction of hypoxia-inducible factor 1alpha in vascular smooth muscle cells." J Biol Chem **275**(35): 26765-26771.
- Roulet, E., I. Fisch, et al. (1998). "Evaluation of computer tools for the prediction of transcription factor binding sites on genomic DNA." In Silico Biol **1**(1): 21-28.
- Ryan, H. E., J. Lo, et al. (1998). "HIF-1 alpha is required for solid tumor formation and embryonic vascularization." EMBO J **17**(11): 3005-3015.
- Salceda, S. and J. Caro (1997). "Hypoxia-inducible factor 1alpha (HIF-1alpha) protein is rapidly degraded by the ubiquitin-proteasome system under normoxic conditions. Its stabilization by hypoxia depends on redox-induced changes." J Biol Chem **272**(36): 22642-22647.
- Sanchez-Elsner, T., L. M. Botella, et al. (2001). "Synergistic cooperation between hypoxia and transforming growth factor-beta pathways on human vascular endothelial growth factor gene expression." J Biol Chem **276**(42): 38527-38535.
- Sang, N., D. P. Stiehl, et al. (2003). "MAPK signaling up-regulates the activity of hypoxia-inducible factors by its effects on p300." J Biol Chem **278**(16): 14013-14019.
- Schneider, T. D., G. D. Stormo, et al. (1986). "Information content of binding sites on nucleotide sequences." J Mol Biol **188**(3): 415-431.
- Semenza, G. L. (1999). "Perspectives on oxygen sensing." Cell **98**(3): 281-284.
- Semenza, G. L. (2003). "Targeting HIF-1 for cancer therapy." Nat Rev Cancer **3**(10): 721-732.
- Semenza, G. L. (2004). "Hydroxylation of HIF-1: oxygen sensing at the molecular level." Physiology (Bethesda) **19**: 176-182.
- Semenza, G. L. and G. L. Wang (1992). "A nuclear factor induced by hypoxia via de novo protein synthesis binds to the human erythropoietin gene enhancer at a site required for transcriptional activation." Mol Cell Biol **12**(12): 5447-5454.
- Shohet, R. V. and J. A. Garcia (2007). "Keeping the engine primed: HIF factors as key regulators of cardiac metabolism and angiogenesis during ischemia." J Mol Med **85**(12): 1309-1315.
- Siepel, A., G. Bejerano, et al. (2005). "Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes." Genome Res **15**(8): 1034-1050.
- Sinha, S., E. van Nimwegen, et al. (2003). "A probabilistic method to detect regulatory modules." Bioinformatics **19 Suppl 1**: i292-301.
- Sodhi, A., S. Montaner, et al. (2001). "MAPK and Akt act cooperatively but independently on hypoxia inducible factor-1alpha in rasV12 upregulation of VEGF." Biochem Biophys Res Commun **287**(1): 292-300.
- Sumbayev, V. V., A. Budde, et al. (2003). "HIF-1 alpha protein as a target for S-nitrosation." FEBS Lett **535**(1-3): 106-112.

## Bibliografía

- Tagle, D. A., B. F. Koop, et al. (1988). "Embryonic epsilon and gamma globin genes of a prosimian primate (*Galago crassicaudatus*). Nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints." *J Mol Biol* **203**(2): 439-455.
- Takahashi, K., K. Koga, et al. (2009). "Macrophage CD74 contributes to MIF-induced pulmonary inflammation." *Respir Res* **10**: 33.
- Takeda, K., H. L. Aguila, et al. (2008). "Regulation of adult erythropoiesis by prolyl hydroxylase domain proteins." *Blood* **111**(6): 3229-3235.
- Takeda, K., V. C. Ho, et al. (2006). "Placental but not heart defects are associated with elevated hypoxia-inducible factor alpha levels in mice lacking prolyl hydroxylase domain protein 2." *Mol Cell Biol* **26**(22): 8336-8346.
- Tang, N., L. Wang, et al. (2004). "Loss of HIF-1alpha in endothelial cells disrupts a hypoxia-driven VEGF autocrine loop necessary for tumorigenesis." *Cancer Cell* **6**(5): 485-495.
- Taylor, M. S. (2001). "Characterization and comparative analysis of the EGLN gene family." *Gene* **275**(1): 125-132.
- Tian, H., R. E. Hammer, et al. (1998). "The hypoxia-responsive transcription factor EPAS1 is essential for catecholamine homeostasis and protection against heart failure during embryonic development." *Genes Dev* **12**(21): 3320-3324.
- Tian, H., S. L. McKnight, et al. (1997). "Endothelial PAS domain protein 1 (EPAS1), a transcription factor selectively expressed in endothelial cells." *Genes Dev* **11**(1): 72-82.
- van Helden, J. (2003). "Regulatory sequence analysis tools." *Nucleic Acids Res* **31**(13): 3593-3596.
- Vigoda, A., L. K. Mamedova, et al. (2003). "Glycogen metabolism in rat heart muscle cultures after hypoxia." *Mol Cell Biochem* **254**(1-2): 311-318.
- Vinores, S. A., W. H. Xiao, et al. (2006). "Implication of the hypoxia response element of the Vegf promoter in mouse models of retinal and choroidal neovascularization, but not retinal vascular development." *J Cell Physiol* **206**(3): 749-758.
- Visel, A., E. M. Rubin, et al. (2009). "Genomic views of distant-acting enhancers." *Nature* **461**(7261): 199-205.
- Webster, K. A. (2007). "Hypoxia: life on the edge." *Antioxid Redox Signal* **9**(9): 1303-1307.
- Wen, W., J. Ding, et al. (2010). "Suppression of cyclin D1 by hypoxia-inducible factor-1 via direct mechanism inhibits the proliferation and 5-fluorouracil-induced apoptosis of A549 cells." *Cancer Res* **70**(5): 2010-2019.
- Wenger, R. H., D. P. Stiehl, et al. (2005). "Integration of oxygen signaling at the consensus HRE." *Sci STKE* **2005**(306): re12.
- Wiesener, M. S., H. Turley, et al. (1998). "Induction of endothelial PAS domain protein-1 by hypoxia: characterization and comparison with hypoxia-inducible factor-1alpha." *Blood* **92**(7): 2260-2268.
- Willam, C., P. H. Maxwell, et al. (2006). "HIF prolyl hydroxylases in the rat; organ distribution and changes in expression following hypoxia and coronary artery ligation." *J Mol Cell Cardiol* **41**(1): 68-77.
- Xia, X., M. E. Lemieux, et al. (2009). "Integrative analysis of HIF binding and transactivation reveals its role in maintaining histone methylation homeostasis." *Proc Natl Acad Sci U S A* **106**(11): 4260-4265.
- Yang, X., M. Guo, et al. (2010). "Deregulation of growth factor, circadian clock, and cell cycle signaling in regenerating hepatocyte RXRalpha-deficient mouse livers." *Am J Pathol* **176**(2): 733-743.
- Yu, A. Y., L. A. Shimoda, et al. (1999). "Impaired physiological responses to chronic hypoxia in mice partially deficient for hypoxia-inducible factor 1alpha." *J Clin Invest* **103**(5): 691-696.
- Zelzer, E., Y. Levy, et al. (1998). "Insulin induces transcription of target genes through the hypoxia-inducible factor HIF-1alpha/ARNT." *EMBO J* **17**(17): 5085-5094.



- Zhang, H., M. Bosch-Marce, et al. (2008). "Mitochondrial autophagy is an HIF-1-dependent adaptive metabolic response to hypoxia." J Biol Chem **283**(16): 10892-10903.
- Zhang, W., T. Tsuchiya, et al. (1999). "Transitional change in interaction between HIF-1 and HNF-4 in response to hypoxia." J Hum Genet **44**(5): 293-299.
- Zhong, H., K. Chiles, et al. (2000). "Modulation of hypoxia-inducible factor 1alpha expression by the epidermal growth factor/phosphatidylinositol 3-kinase/PTEN/AKT/FRAP pathway in human prostate cancer cells: implications for tumor angiogenesis and therapeutics." Cancer Res **60**(6): 1541-1545.

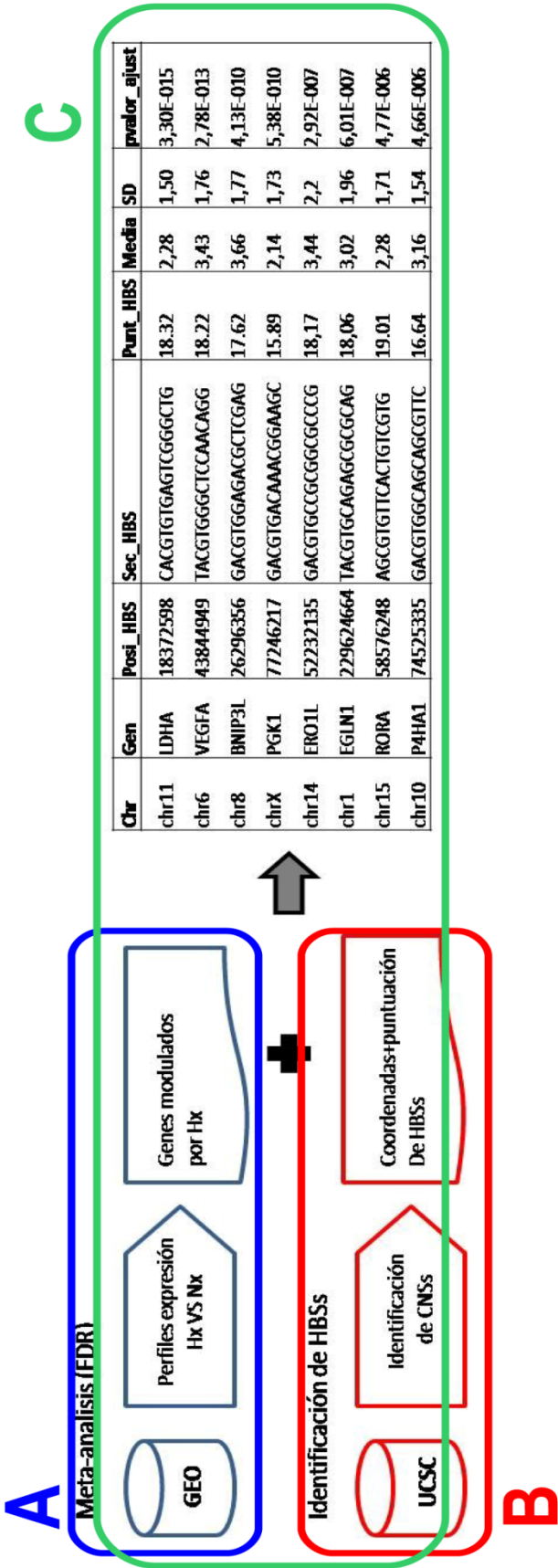


***ANEXOS***

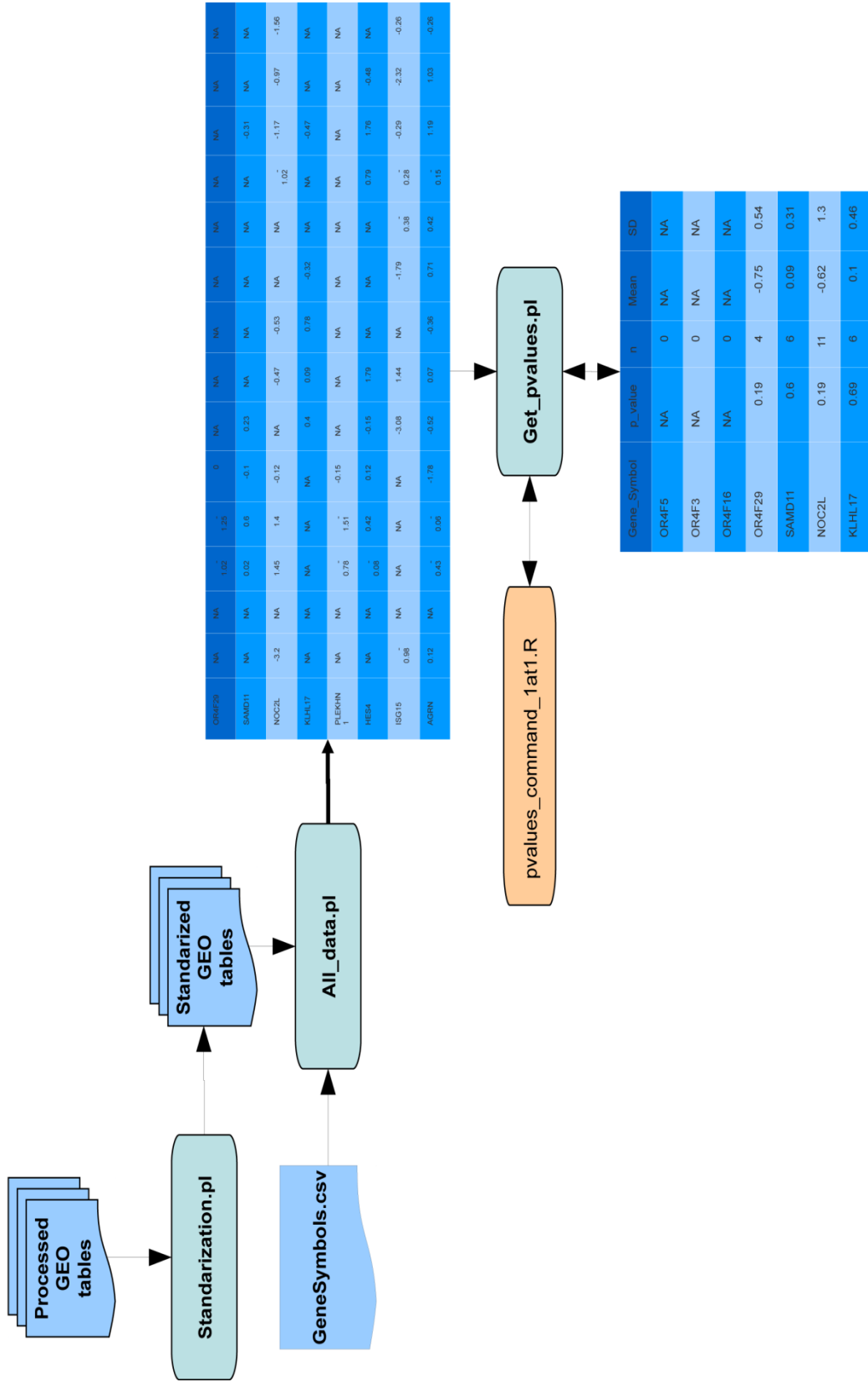


IX.1 ANEXO I

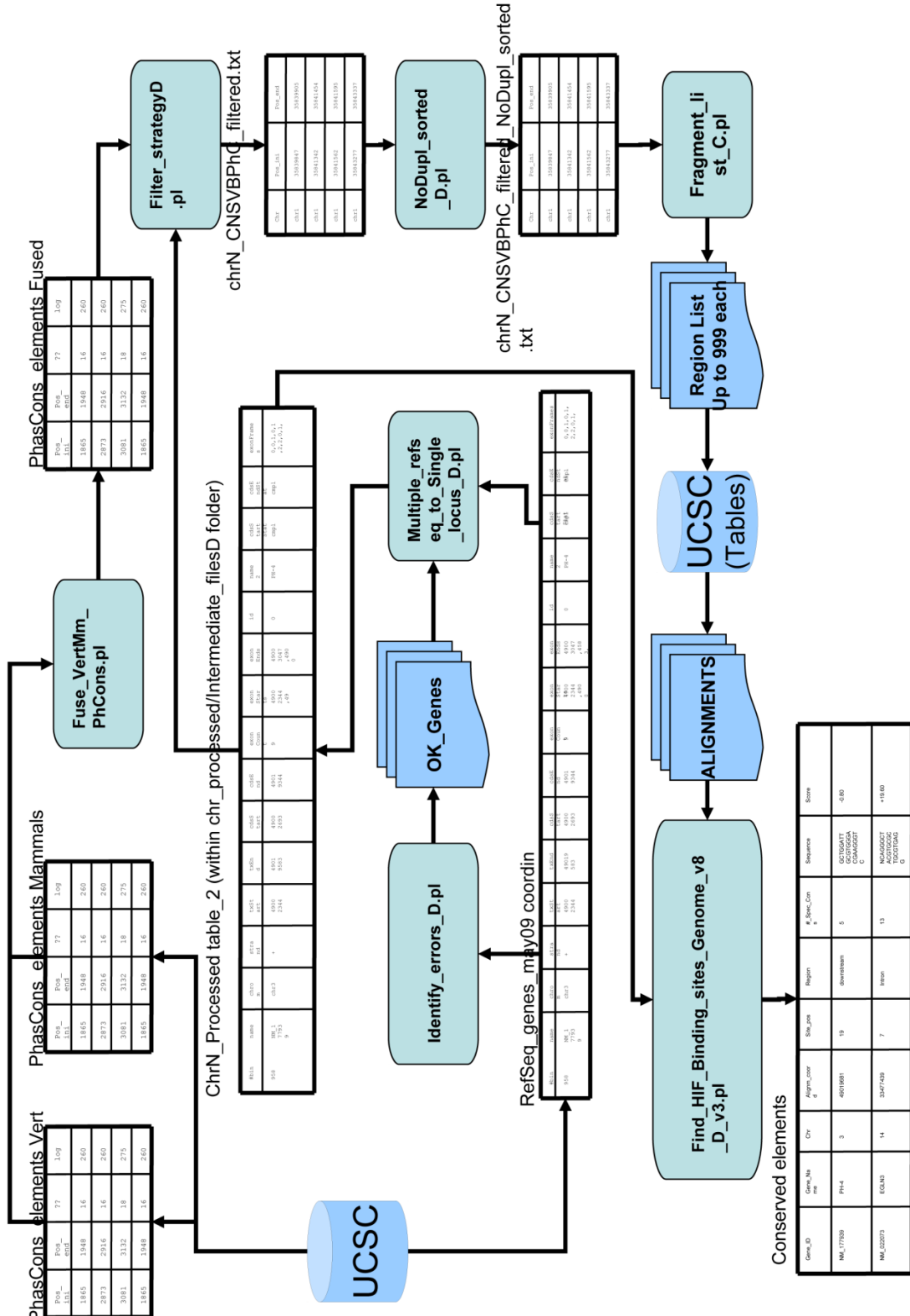
**Anexo I: Esquema de la Estrategia Bioinformática.** (A) Meta-análisis de los perfiles de expresión de GEO. (B) Identificación de sitios de unión de HIF (HBSs) (C) Estrategia Bioinformática, listado de potenciales dianas de HIF (D) Descripción de los programas ("scripts") utilizados en el desarrollo de la estrategia. (E) Programas ("scripts")



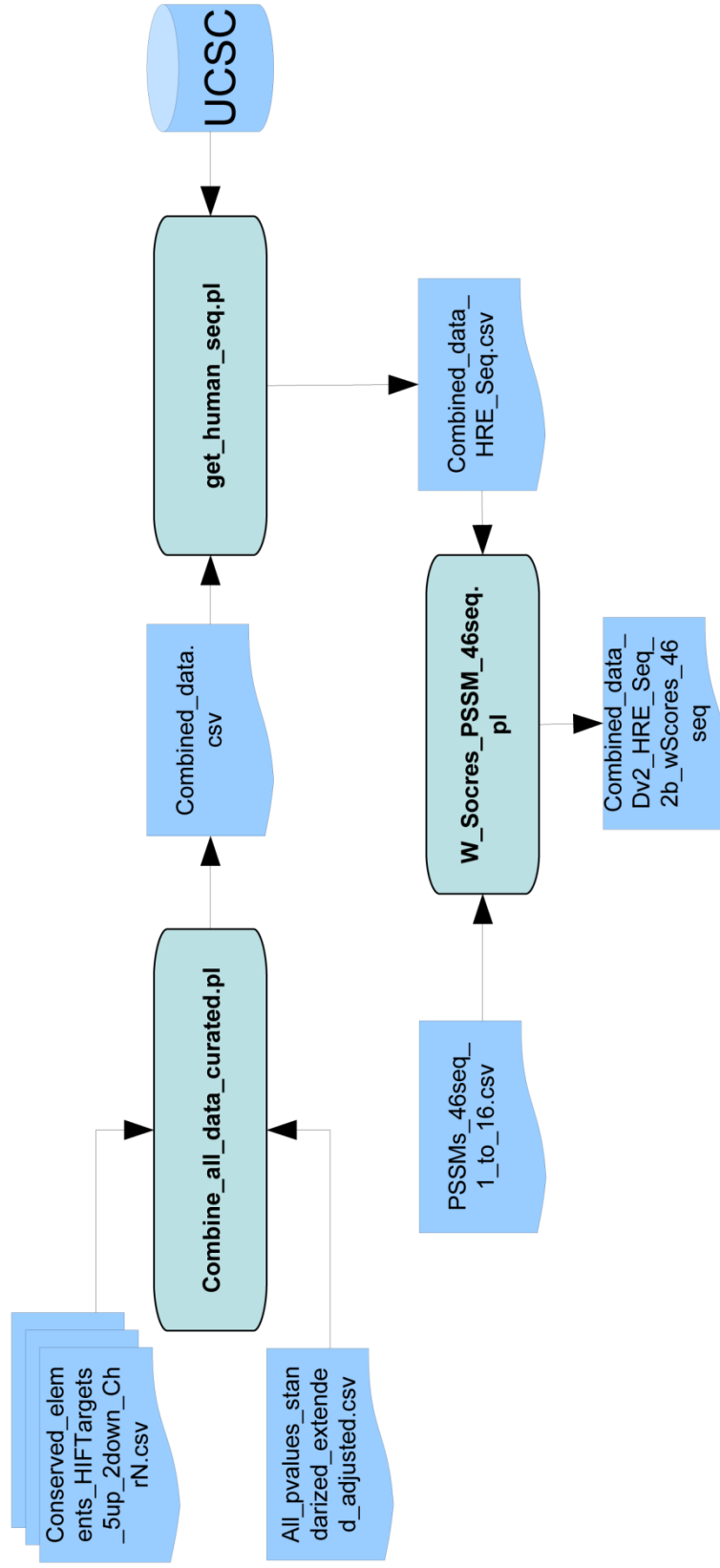
# A Meta-analysis de perfiles de expresión génica en condiciones de hipoxia



# B Identificación de sitios de unión de HIF *in silico*



## C Intersección del meta-análisis y los sitios potenciales de unión de HIF





## E

Nombre	Parte	Descripción	Entrada	Salida
<i>Standardization.pl</i>	A	Estandariza (resta media de valores log-ratio Hx/Nx de la tabla y divide por la SD) cada una de las 19 tablas GEO	Tablas GEO	Tablas GEO estandarizadas
<i>All_data.pl</i>	A	Recopila el log-ratio Hx/Nx estandarizado de todas las sondas para un locus génico dado	Tablas GEO estandarizadas Lista de "gene symbols"	Tabla con todos los valores log-ratio Hx/Nx estandarizados de todos los genes
<i>Get_pvalues.pl</i>	A	Calcula la media y SD de todos los log-ratio Hx/Nx estandarizados para cada locus y lo compara (test de student simple) con la hipótesis nula (valor 0)	Tabla con todos los valores log-ratio Hx/Nx estandarizados de todos los genes	Tabla con todos la media, Sd y pvalor de los log-ratio Hx/Nx estandarizados de todos los genes
<i>Identify_errors_D.pl</i>	B	Identifica genes duplicados, genes dentro de intrones, y otras regiones que no se pueden analizar fácilmente y los elimina	Tablas conteniendo la información de todos los mRNA "curados" (RefSeq) del genoma. Una tabla por cada cromosoma	Listado conteniendo los "gene symbol" de todos los genes del genoma exceptuando los contenidos en regiones "problemáticas" Una tabla por cada cromosoma
<i>Multiple_refseq_to_Single_locus_D.pl</i>	B	Genera tablas conteniendo la información de todos los mRNA "curados" (RefSeq) del genoma localizados fuera de regiones problemáticas) e incluyendo las coordenadas de la regiones no codificantes de cada locus (incluyendo regiones upstream del TSS). Una tabla por cada cromosoma	Tablas conteniendo la información de todos los mRNA "curados" (RefSeq) del genoma. Una tabla por cada cromosoma. Listado de "gene symbols" de genes a procesar (fuera de regiones problemáticas)	Tablas conteniendo la información de todos los mRNA "curados" (RefSeq) del genoma (exceptuando todos los contenidos en regiones "problemáticas") incluyendo las coordenadas de todas las regiones no codificantes Una tabla por cada cromosoma
<i>Fuse_VerthMm_PhCons.pl</i>	B	Compara elementos PhasCons de vertebrados y de Mamíferos y los fusiona	Tabla elementos PhasCons vertebrados Tabla elementos PhasCons Mamíferos	Tabla elementos PhasCons fusionados
<i>Filter_strategyD.pl</i>	B	Identifica elementos PhasCons (fused) contenidos en regiones no codificantes de los locus génicos. Fusiona elementos adyacentes siempre que en la fusión resultante, al menos, el 50% de las bases pertenezcan a un PhasCons	Tablas conteniendo la información de todos los mRNA "curados" (RefSeq) del genoma (exceptuando todos los contenidos en regiones "problemáticas") incluyendo las coordenadas de todas las regiones no codificantes Una tabla por cada cromosoma	Tabla coordenadas regiones No codificantes conservadas (CNSs)
<i>NoDup_sorted_D.pl</i>	B	Comprueba que no haya duplicados de CNSs	Tabla coordenadas regiones No codificantes conservadas (CNSs)	Tabla coordenadas regiones No codificantes conservadas (CNSs) no redundante
<i>Fragment_list_C.pl</i>	B	Fragmenta los listados de coordenadas en bloques de 1000 para su uso en el servidor UCSC	Tabla coordenadas regiones No codificantes conservadas (CNSs) no redundante	Tablas coordenadas regiones No codificantes conservadas (CNSs) no redundantes en bloques de 1000 regiones

Nombre	Parte	Descripción	Entrada	Salida
<i>Find_HIF_Binding_sites_Genome_v8_D_v3.pl</i>	B	Rastrea los alineamientos múltiples (descargados del navegador genómico del UCSC) correspondientes a las coordenadas de las CNSs en busca de motivos RCGTG. Para cada motivo cuenta el número de especies en las que está conservado y solo considera los conservados en Mm Y (Rn O Cf O Ee O La O Md) Y (Rm O Pt). Por último da una puntuación a cada motivo en base a la PSSM.	Alineamientos múltiples correspondientes a las CNSs Tablas de RefSeq conteniendo coordenadas de regiones no codificantes	Tabla con todos los locus génicos estudiados, conteniendo información sobre la presencia de motivos RCGTG, sus coordenadas cromosómicas, conservación del motivo (#especies en el alineamiento que conservan el motivo), puntuación y máxima puntuación del locus
<i>Combine_all_data_curated.pl</i>	C	Combina en una sola tabla la información sobre inducción de cada gen (media, SD y p-valor) y los potenciales sitios de unión de HIF en ese locus	Tabla con todos los locus génicos estudiados, conteniendo información sobre la presencia de motivos RCGTG, sus coordenadas cromosómicas, conservación del motivo (#especies en el alineamiento que conservan el motivo), puntuación y máxima puntuación del locus. Tabla con todos la media,SD y pvalor de los log-ratio Hx/Nx estandarizados de todos los genes	Tabla con todos los locus génicos estudiados, conteniendo información sobre la presencia de motivos RCGTG, sus coordenadas cromosómicas, conservación del motivo (#especies en el alineamiento que conservan el motivo), puntuación y máxima puntuación del locus y la median ,SD y pvalor de los log-ratio Hx/Nx estandarizados de todos los genes
<i>get_human_seq_v4.pl</i>	C	Obtiene la secuencia completa de los 33bp conteniendo cada uno de los potenciales HBS del genoma humano	Tabla salida de Combine_all_data Conexión a UCSC	Incluye en la Tabla de entrada la secuencia completa.
<i>W_Scores_PSSM_46seq.pl</i>	C	Puntúa los HBS de acuerdo a la PSSM utilizando la secuencia completa conteniendo el sitio.	Tabla salida de get_human_seq PSSM (PSSMs_46seq_1_to_16.csv)	Incluye en la tabla de entrada las puntuaciones

**(E) SCRIPTS****Standarization.pl.**

```

#!/usr/bin/perl -w
use strict;
my ($i,$ff,$Sd,$Mean);
my @files=<Tables_extended/G*>;
print @files;
foreach $ff (@files){
    $ff=~s/(\w+\/)//;
    print $ff,"\n";
    open (DATA,"Tables_extended/$ff") or die "file $ff not found\n";
    open (RES, ">Standarized_extended/$ff") or die "Cannot create file
$ff\n";
    while (<DATA>){
        chomp $_;$_=~s/"//g;my @tmp=split "\t",$_;
        if ($_~/GEO_ID|Cell|Treatment|Symbol/){print RES
$tmp[0],"\t",$tmp[1],"\t",$tmp[2],"\t",$tmp[3],"\n";}
        elsif ($_~/SD\sSD/){
            $Sd=$tmp[3];
            print RES
$tmp[0],"\t",$tmp[1],"\t",$tmp[2],"\t",$tmp[3],"\n";
        }
        elsif ($_~/Mean/){
            $Mean=$tmp[3];
            print RES
$tmp[0],"\t",$tmp[1],"\t",$tmp[2],"\t",$tmp[3],"\n";
        }
        else{
            print RES $tmp[0],"\t",$tmp[1],"\t",$tmp[2],"\t",(($tmp[3]-
$Mean)/$Sd),"\n";
        }
    }
}
exit;

```

**All data.pl.**

```

#!/usr/bin/perl -w
use strict;
my @Gsym;my $i;my @tmp; my @files;
open (DATA, "GeneSymbols.csv") or die "file not found\n";
    while(<DATA>){chomp $_;$_=~s/"//g;push @Gsym, (split "\t",$_)[0];print
$Gsym[-1],"\n";}
close (DATA);
chdir ("./Standarized_extended");@files=glob ("*");chdir
("../pvalues_standarized_extended");
print scalar(@files),"\n";
for ($i=0;$i<scalar(@files);$i++){print $files[$i],"\n";}
open (RES, ">All_data.csv") or die "Unable to write file\n";
#for ($i=0;$i<scalar(@Gsym);$i++){
for ($i=0;$i<scalar(@Gsym);$i++){
    print RES $Gsym[$i];
    print "$i $Gsym[$i]\n";
    foreach my $file (@files){
        #open (DATA, "./Tables/$file.csv") or die "file $file not
found\n";
        open (DATA, "../Standarized_extended/$file") or die "file $file
not found\n";
        while (<DATA>){
            $_=~s/"//g;
            chomp $_;
            @tmp=split "\t",$_;

```

## Anexol

```
        if (uc($tmp[0]) eq uc($Gsym[$i])){print RES "\t",$tmp[-1];}
    }
    close(DATA);
}
print RES "\n";
}
close(RES);
exit;
```

## Get values.pl

```
#!/usr/bin/perl -w
use strict;
my $i;my $j=0;my @tmp;
open (DATA, "./pvalues_standarized_extended/All_data.csv") or die "file not
found\n";
open (RES1,
">./pvalues_standarized_extended/All_pvalues_standarized_extended.csv") or die
"Unable to create file\n";
print RES1 "Gene_Symbol\tp_value\tMean\tSD\n";
while(<DATA>){
    $j++;
    chomp $_;$_ =~ s/"//;@tmp=split "\t",$_;
    if (scalar(@tmp)>2){
        open (RES2, ">gene_data.tmp");
        for ($i=0;$i<scalar(@tmp);$i++){print RES2 $tmp[$i],"\n";}
        close (RES2);
        system ("Rscript pvalues_command_lat1.R gene_data.tmp");
        open (DATA2, "./pvalues_standarized_extended/pval.csv");my
$pv=<DATA2>;chomp $pv;close(DATA2);
        my $mm=&MX(\@tmp);
        print RES1
$tmp[0],"\t",$pv,"\t",scalar(@tmp),"\t",$mm,"\t",&Sd(\@tmp,$mm),"\n";
        print $j,"\t",$tmp[0],"\t",$pv,"\t", (scalar(@tmp)-
1),"\t",$mm,"\t",&Sd(\@tmp,$mm),"\n";
    }
    else{
        print RES1 $tmp[0],"\tNA\t", (scalar(@tmp)-1),"\tNA\tNA\n";
        print $tmp[0],"\tNA\t", (scalar(@tmp)-1),"\tNA\tNA\n";
    }
}
close(RES1);
close(DATA);
exit;

sub MX{
    my ($ad)=@_;
    my $i;my $n=0;
    for ($i=1;$i<scalar(@$ad);$i++){ $n=$n+$ad[$i];};
    return ($n/(scalar(@$ad)-1))
}

sub Sd{
    my ($ad,$m)=@_;
    my $i;my $n=0;
    for ($i=1;$i<scalar(@$ad);$i++){ $n=$n+ (($m-$ad[$i])**2);};
    return (( $n/(scalar(@$ad)-1) )**0.5)
}

}
```

Identify errors.pl

```

#!/usr/bin/perl -w
#This script identifies genes containing other genes as well as duplicated
genes that contains other genes in the region between them. All these genes
are eliminated
# May2009
# Luis del Peso

my $chr=$ARGV[0] or die "No chr passed as argument\n";
my %GS_l;my %GS_r; #outer left (%GS_l) and right (%GS_r) limits for a given
gene symbol (key of each hash)
print "Processing chr$chr\n";
my $file="../Genome_files/RefSeq_genes_may09/chr".$chr.".txt";
open (DATA, $file) or die "Cannot open DATA\n";
while (my $line=<DATA>){
    if($line=~/\bin/){next;}
    chomp $line;$line=~s/"//g;my @tmp=split "\t",$line;
    unless ($tmp[4] and $tmp[5]){next;}
    unless ($tmp[4]<$tmp[5]){next;}
    if (exists($GS_l{$tmp[12]})){
        if ($tmp[4]<$GS_l{$tmp[12]}){$GS_l{$tmp[12]}=$tmp[4];}
        if ($tmp[5]>$GS_r{$tmp[12]}){$GS_r{$tmp[12]}=$tmp[5];}
    }
    else{$GS_l{$tmp[12]}=$tmp[4];$GS_r{$tmp[12]}=$tmp[5];}
    #print $tmp[12],"\t",$GS_l{$tmp[12]},"\t",$GS_r{$tmp[12]},"\n";
}
close(DATA);
my @Gs=keys(%GS_l);my %BadGS;
# open(RESx, ">Genes_OK_may09/control_$chr.csv");foreach my $tt(@Gs){print
RESx $tt,"\t",$GS_l{$tt},"\\t",$GS_r{$tt},"\\n";}close(RESx);
open (RES, ">Genes_OK_may09/errors_chr".$chr.".csv") or die "unable to
save RES\n";
open (RES2, ">Genes_OK_may09/OK_genes_chr".$chr.".csv") or die "unable
to save RES2\n";
for ($i=0;$i<scalar(@Gs);$i++){
    my $rk=0;
    for ($j=0;$j<scalar(@Gs);$j++){
        if ($i==$j){next;}
        my $Li=$GS_l{$Gs[$i]};my $Lj=$GS_l{$Gs[$j]};my
        $Ri=$GS_r{$Gs[$i]};my $Rj=$GS_r{$Gs[$j]};
        #print
        $Gs[$i],"\t",$Li,"\\t",$Ri,"\\t",$Gs[$j],"\t",$Lj,"\\t",$Rj,"\\n";
        if ($Li==$Lj and $Ri==$Rj){next;}
        if (($Li<$Lj and $Lj<$Ri)or($Lj<$Li and $Li<$Rj)){
            #print "problem found\n";
            $rk=1;
            if
            (exists($BadGS{$Gs[$i]}){$BadGS{$Gs[$i]}=$BadGS{$Gs[$i]}+1;}else{$BadGS{$Gs[$
            i]}=1;print RES $Gs[$i],"\\n";}
            if
            (exists($BadGS{$Gs[$j]}){$BadGS{$Gs[$j]}=$BadGS{$Gs[$j]}+1;}else{$BadGS{$Gs[$
            j]}=1;print RES $Gs[$j],"\\n";}
        }
    }
    if ($rk<1){print RES2 $Gs[$i],"\\n";}
}
close(RES);
close (RES2);

exit;

```

Multiple Ref Seq to Single Locus.pl.

```

#!/usr/bin/perl -w
#This program parses a list of RefSeq mRNA coordinates and transforms it into
list of locus coordinates.
#Requires a RefSeq list of mRNAs (from UCSC) as input
# May 2009
# Luis del Peso
use strict;
my ($n,$line,$i);
my %Gene_symbol;
my %GS_OK;
my $GeneTable_name;
if (@ARGV){$GeneTable_name=$ARGV[0];}else{$GeneTable_name="";}

# Read list of OK genes
open (DATA, "Genes_OK_may09/OK_genes_chr".$GeneTable_name.".csv")or die
"Genes_OK not found\n";
while (<DATA>){chomp $_;$_ =~s/"//g;$GS_OK{$_}=$_;}
close (DATA);
my
$file="../Genome_files/RefSeq_genes_may09/chr".$GeneTable_name.".txt";#(path
and)file name containing gene table

#Identify GeneSymbols,counts number of refseq per GeneSymbol and process Exons
in each refseq to split CDS and UTRs
unless(open (RES, ">RefSeq_processed_D/chr".$GeneTable_name.".csv")){die
"Cannot open $GeneTable_name\n";}
print RES
"bin\tname\tchrom\tstrand\ttxStart\ttxEnd\tcdsStart\tcdsEnd\texonCount\texonSt
arts\texonEnds\texonType\tname2\n";
open(DATA, $file) or die "file $file not found";
while ($line=<DATA>){
  if ($line=~/#bin/){next;}
  chomp $line;
  $line=~s/"//g;
  my @temp=split("\t",$line);
  unless ($GS_OK{$temp[12]}){next;}
  unless ($temp[7]>$temp[6]){next;}
  if($Gene_symbol{$temp[12]}){$Gene_symbol{$temp[12]}=$Gene_symbol{$temp[1
2]}+1;}else{$Gene_symbol{$temp[12]}=1;}
  my @Genes=@temp;
  &Get_info(\@Genes);
  for ($i=0;$i<13;$i++){print RES $Genes[$i],"\t";}
  print RES "\n";
}
#foreach (keys(%Gene_symbol)){print $_,"\t", $Gene_symbol{$_},"\n";}
close (DATA);
close (RES);

#Process each Locus: merges all exons for a given locus into a single row
open (RES, ">Processed/chr".$GeneTable_name."_Processed.csv")or die "Unable to
write file\n";
print RES
"bin\tname\tchrom\tstrand\ttxStart\ttxEnd\tcdsStart\tcdsEnd\texonCount\texonSt
arts\texonEnds\texonType\tname2\tLeftGeneEnd\tRightGeneEnd\n";
$file="../RefSeq_processed_D/chr".$GeneTable_name.".csv";
foreach my $GSymb(keys(%Gene_symbol)){
  my @Genes=();
  # print $GSymb,"\n";
  unless(open(DATA, $file)){die "file $file not found";}
  while ($line=<DATA>){
    if ($line=~/#bin/){next;}
    chomp $line;
    $line=~s/"//g;
    my @temp=split("\t",$line);
    if($temp[12] eq $GSymb){

```

```

        if
($Genes[0]){&Merge_info(\@Genes,\@temp);}else{@Genes=@temp;}
    }
    }
    close(DATA);
    &Gene_limits(\@Genes);
    for ($i=0;$i<15;$i++){print RES $Genes[$i],"\t";}
    print RES "\n";
}
close(RES);
exit;

sub Get_info{
    my ($rG)=@_;
    my ($i,$EI,$EE,$EL);
    #Search for exons containing UTR and CDS and divide them in two
consecutive artificial exons
    $$rG[11]="";$EI="";$EE="";$EL="";
    my @ei=split ",", $$rG[9];
    my @ee=split ",", $$rG[10];
    for ($i=0;$i<scalar(@ei);$i++){
        if ($ee[$i]<$$rG[6]){#exon before CDS => UTR
            $EI=$EI.$ei[$i].",";
            $EE=$EE.$ee[$i].",";
            if ($$rG[3]=~/\+/){$EL=$EL."5UTR,";}else{$EL=$EL."3UTR,";}
        }
        elsif ($ei[$i]>$$rG[7]){#exon after CDS => UTR
            $EI=$EI.$ei[$i].",";
            $EE=$EE.$ee[$i].",";
            if ($$rG[3]=~/\+/){$EL=$EL."3UTR,";}else{$EL=$EL."5UTR,";}
        }
        elsif ($ei[$i]>=$$rG[6] and $ee[$i]<=$$rG[7]){#exon after start
and before end => CDS
            $EI=$EI.$ei[$i].",";
            $EE=$EE.$ee[$i].",";
            $EL=$EL."CDS,";
        }
        elsif ($ei[$i]<$$rG[6] and $ee[$i]>$$rG[6] and
$ee[$i]<=$$rG[7]){#exon contains start CDS => Partly CDS and partly UTR --
>divide into two exons
            $EI=$EI.$ei[$i].",".($$rG[6]+1).",";
            $EE=$EE.($$rG[6]-1).",".$ee[$i].",";
            if
($$rG[3]=~/\+/){$EL=$EL."5UTR,CDS,";}else{$EL=$EL."3UTR,CDS,";}
            $$rG[8]=$$rG[8]+1;
        }
        elsif ($ei[$i]<$$rG[7] and $ee[$i]>$$rG[7] and
$ei[$i]>=$$rG[6]){#exon contains Stop CDS => Partly CDS and partly UTR --
>divide into two exons
            $EI=$EI.$ei[$i].",".($$rG[7]+1).",";
            $EE=$EE.($$rG[7]-1).",".$ee[$i].",";
            if
($$rG[3]=~/\+/){$EL=$EL."CDS,3UTR,";}else{$EL=$EL."CDS,5UTR,";}
            $$rG[8]=$$rG[8]+1;
        }
        elsif ($ei[$i]<$$rG[6] and $ee[$i]>$$rG[6] and $ei[$i]<$$rG[7]
and $ee[$i]>$$rG[7]){#exon contains Start AND Stop CDS => Partly CDS and
partly UTR -->divide into three exons
            $EI=$EI.$ei[$i].",".($$rG[6]+1).",".($$rG[7]+1).",";
            $EE=$EE.($$rG[6]-1).",".($$rG[7]-1).",".$ee[$i].",";
            if
($$rG[3]=~/\+/){$EL=$EL."5UTR,CDS,3UTR,";}else{$EL=$EL."3UTR,CDS,5UTR,";}
            $$rG[8]=$$rG[8]+2;
        }
    }
    $$rG[9]=$EI;$rG[10]=$EE;$rG[11]=$EL;
}

```

```

sub Merge_info{
  my ($rG,$rT)=@_;
  #print "\t Sub Merge\n$$rG[9]\n$$rT[9]\n";
  unless ($$rG[4]=~/ $$rT[4]/) {$$rG[4]=$$rG[4].",".$$rT[4];}
  unless ($$rG[5]=~/ $$rT[5]/) {$$rG[5]=$$rG[5].",".$$rT[5];}
  my @tmp1i=split ",",$rG[9];
  my @tmp1e=split ",",$rG[10];
  my @tmp1t=split ",",$rG[11];
  my @tmp2i=split ",",$rT[9];
  my @tmp2e=split ",",$rT[10];
  my @tmp2t=split ",",$rT[11];
  my $n1=0;my $n2=0;my $d1=0;my $d2=0;my $cond=0; my $EI="";my $EE="";my
  $ET=""; my $EN="";
  while ($cond<1){
    if
    ($d1){$EI=$EI.$tmp2i[$n2].",";$EE=$EE.$tmp2e[$n2].",";$ET=$ET.$tmp2t[$n2].",";
    $n2++;}

    elsif($d2){$EI=$EI.$tmp1i[$n1].",";$EE=$EE.$tmp1e[$n1].",";$ET=$ET.$tmp1
    t[$n1].",";$n1++;}
    else{#Simple splicing scheme
      if ($tmp1i[$n1]==$tmp2i[$n2] and
      $tmp1e[$n1]==$tmp2e[$n2]){#Same Exons
        $EI=$EI.$tmp1i[$n1].",";$EE=$EE.$tmp1e[$n1].",";
        if ($tmp2t[$n2]=~/CDS/ or
        $tmp1t[$n1]=~/CDS/){$ET=$ET."CDS,";}else{$ET=$ET."UTR,";}
        $n1++;$n2++;
      }
      elsif ($tmp2e[$n2]<$tmp1i[$n1]){#Exon 1 after Exon 2
        $EI=$EI.$tmp2i[$n2].",";$EE=$EE.$tmp2e[$n2].",";$ET=$ET.$tmp2t[$n2].",";
        $n2++;
      }
      elsif ($tmp2i[$n2]>$tmp1e[$n1]){#Exon 2 after Exon 1
        $EI=$EI.$tmp1i[$n1].",";$EE=$EE.$tmp1e[$n1].",";$ET=$ET.$tmp1t[$n1].",";
        $n1++;
      }
      else{#Exon(s) contained in other Exon
        if ($tmp1i[$n1]<$tmp2i[$n2]){
          if (($tmp2i[$n2]-$tmp1i[$n1])>3){
            $EI=$EI.$tmp1i[$n1].",";$EE=$EE.($tmp2i[$n2]-1).",";
            if ($tmp2t[$n2]=~/CDS/ or
            $tmp1t[$n1]=~/CDS/){$ET=$ET."CDS,";}else{$ET=$ET."UTR,";}
          }
          $tmp1i[$n1]=$tmp2i[$n2];
        }
        elsif ($tmp1i[$n1]>$tmp2i[$n2]){
          if (($tmp1i[$n1]-$tmp2i[$n2])>3){
            $EI=$EI.$tmp2i[$n2].",";$EE=$EE.($tmp1i[$n1]-1).",";
            if ($tmp2t[$n2]=~/CDS/ or
            $tmp1t[$n1]=~/CDS/){$ET=$ET."CDS,";}else{$ET=$ET."UTR,";}
          }
          $tmp2i[$n2]=$tmp1i[$n1];
        }
        elsif ($tmp1e[$n1]<$tmp2e[$n2]){
          $EI=$EI.$tmp1i[$n1].",";$EE=$EE.$tmp1e[$n1].",";
          if ($tmp2t[$n2]=~/CDS/ or
          $tmp1t[$n1]=~/CDS/){$ET=$ET."CDS,";}else{$ET=$ET."UTR,";}
          $tmp2i[$n2]=$tmp1e[$n1]+1;
          $n1++;
          unless ($tmp2e[$n2]-$tmp2i[$n2]>3){$n2++;}
        }
        elsif ($tmp1e[$n1]>$tmp2e[$n2]){

```



```

    $EI=$EI.$tmp1i[$n1].",";$EE=$EE.$tmp2e[$n2].",";
        if ($tmp2t[$n2]=~/CDS/ or
$tmp1t[$n1]=~/CDS/) {$ET=$ET."CDS,";}else{$ET=$ET."UTR,";}
        $tmp1i[$n1]=$tmp2e[$n2]+1;
        $n2++;
        unless ($tmp1e[$n1]-$tmp1i[$n1]>3) {$n1++;}
    }
    else{die "unknown\n";}
}

}

if ($n1==scalar(@tmp1i)) {$d1=1;}
if ($n2==scalar(@tmp2i)) {$d2=1;}
if ($d1 and $d2) {$cond=1;}
#print "**$EI\n";sleep 1;
}
}
$RrG[9]=$EI;$RrG[10]=$EE;$RrG[11]=$ET;
if ($EN) {$RrG[13]=$EN;}else{$RrG[13]="OK";}
if (exists ($RrG[1])) {$RrG[1]=$RrG[1]."|".$RrT[1];}else{$RrG[1]=$RrT[1];}
}

sub Gene_limits{
    my ($RrG,$i)=@_;
    #print "\t Gene Limits\n";
    my @tmpi=split ",",,$RrG[4];
    my @tmpe=split ",",,$RrG[5];
    my $min_i=1000000000; my $max_e=0;
    for ($i=0;$i<scalar(@tmpi);$i++){
        if ($tmpi[$i]<$min_i) {$min_i=$tmpi[$i];}
    }
    for ($i=0;$i<scalar(@tmpe);$i++){
        if ($tmpe[$i]>$max_e) {$max_e=$tmpe[$i];}
    }
    $RrG[13]=$min_i;
    $RrG[14]=$max_e;
}
}

```

### Fuse VertMm PhCons.pl

```

#!/usr/bin/perl -w
# This program generates a table of coordinates for regions contained within
vertebrate OR Mammal phascons elements
# Jun 2009
# Luis del Peso

my $Chr_name=$ARGV[0] or die "You must indicate chromosome number as an
argument\n";
my
$fV="PhasCons/PhasCons_MostConserv_Vertebrates/UCSC_Ch".$Chr_name."_sorted.txt";
my
$fM="PhasCons/PhasCons_MostConserv_Mammals/UCSC_Ch".$Chr_name."_sorted.txt";
my (@V_i,@V_e,@M_i,@M_e);my ($i,$j);
open (DATA, $fV) or die "file not found\n";while (<DATA>){chomp;$_=~/s//;/g;my
@tmp=split;push @V_i,$tmp[0];push @V_e,$tmp[1];}close (DATA);
open (DATA, $fM) or die "file not found\n";while (<DATA>){chomp;$_=~/s//;/g;my
@tmp=split;push @M_i,$tmp[0];push @M_e,$tmp[1];}close (DATA);
#print $V_i[0],"\t",$V_e[0],"\t",$M_i[0],"\t",$M_e[0],"\n";print $V_i[-
1],"\t",$V_e[-1],"\t",$M_i[-1],"\t",$M_e[-1],"\n";
my ($fin,$tmp_i,$tmp_e,$tt,$org);
$tmp_i=0;$tmp_e=0;$tt=0;
if ($M_e[-1]>=$V_e[-1]) {$fin=$M_e[-1];}else {$fin=$V_e[-1];}
my $ini_time=time;

```

```

open (RES, ">PhasCons/Fused_PhCons_chr".$Chr_name.".csv") or die "Unable to
save file\n";
until($tt==$fin){
  for ($i=0;$i<scalar(@V_i);$i++){
    if ($V_i[$i]>$tt){$tmp_i=$V_i[$i];$tmp_e=$V_e[$i];$org="V";last;}
  }
  for ($i=0;$i<scalar(@M_i);$i++){
    if ($M_i[$i]>$tt and
$M_i[$i]<$tmp_i){$tmp_i=$M_i[$i];$tmp_e=$M_e[$i];$org="M";}
    if ($M_i[$i]>$tmp_i){last;}
  }
  # print "send: $tmp_i\t$tmp_e\t";
  my $tmp_e=&Overlapping_ends($tmp_e,\@V_i,\@V_e,\@M_i,\@M_e,$org);
  # print "get: $tmp_i\t$tmp_e\n";
  print RES $tmp_i,"\t",$tmp_e,"\n";
  $tt=$tmp_e;
  # print "sending V\t$tmp_i\t$tmp_e\t$V_i[0]\t$V_e[0]\n";
  &Remove_coordin($tmp_i,$tmp_e,\@V_i,\@V_e);print "BACK\t$V_i[0]\t$V_e[0]\n";
  # print "sending M\t$tmp_i\t$tmp_e\t$M_i[0]\t$M_e[0]\n";
  &Remove_coordin($tmp_i,$tmp_e,\@M_i,\@M_e);print "BACK\t$M_i[0]\t$M_e[0]\n";
  if(@V_i){&Remove_coordin($tmp_i,$tmp_e,\@V_i,\@V_e);}
  if(@M_i){&Remove_coordin($tmp_i,$tmp_e,\@M_i,\@M_e);}
}
close(RES);
print "Elapsed time: ",time-$ini_time,"\n";
exit;

sub Overlapping_ends{
  my ($te,$rVi,$rVe,$rMi,$rMe,$ro)=@_;
  # print "$te\t$ro\n";
  my $k;
  if ($ro eq "M"){
    for ($k=0;$k<scalar(@$rVi);$k++){
      if ($$rVi[$k]<=$te and
$te<$$rVe[$k]){ $te=&Overlapping_ends($$rVe[$k],$rVi,$rVe,$rMi,$rMe,"V");}
    # if ($$rVi[$k]<=$te and $te<$$rVe[$k]){return ($$rVe[$k]);}
      if ($$rVi[$k]>$te){last;}
    }
    return ($te);
  }
  if ($ro eq "V"){
    for ($k=0;$k<scalar(@$rMi);$k++){
      if ($$rMi[$k]<=$te and
$te<$$rMe[$k]){ $te=&Overlapping_ends($$rMe[$k],$rVi,$rVe,$rMi,$rMe,"M");}
    # if ($$rMi[$k]<=$te and $te<$$rMe[$k]){return ($$rMe[$k]);}
      if ($$rMi[$k]>$te){last;}
    }
    return ($te);
  }
}

sub Remove_coordin{
  my ($ti,$te,$rI,$rE)=@_;
  while ($$rE[0]<=$te){
  # print "Sub\t$$rI[0]\t$$rE[0]\n";
  shift (@$rI);shift (@$rE);
  unless($$rE[-1]){last;}
}
}

```

Filter StrategyD.pl

```

#! usr/bin/perl -w
# This script identifies PhasCons elements within non coding regions of each
gene in the list generated by Multiple_refseq_to_Single_locus_D.pl
# It also identifies PhasCons within $Up bases upstream the transcription
start site and $Down bases downstream the end of the gene
# Adjacent PhasCons regions are fused if the resulting region contains at
least xx% (min_coverage) of conserved bases.
# It expects gene list to process (usually a chromosome in the form: chr1) and
PhasCons element table (either mammals of (more usually) vertebrates) as
arguments
# The output is a list of coordinates corresponding to the start and end of
the CNSs (Conserved non coding regions)
# June 2009
# Luis del Peso

use strict;

my $Up=5000; #number of bases upstream transcription start site
my $Down=1000; #number of bases downstream last exon
my $min_coverage=0.5; #minimal percentage of conserved bases to fuse
contiguous PhasCons elements
my $min_l=5; #minimum region length to be saved
my $Extra_bp=0; #number of bases to extend alignment

my (@Gene_name,@Gene_ini,@Gene_end);
my ($i,$j,@tmp);
my $Chr_name;
my $PhasCons_folder;
my $suffix;
if (@ARGV>1){
    $Chr_name=$ARGV[0];
    if ($ARGV[1]=~/mamm|Mamm|MAMM/){
        $PhasCons_folder='PhasCons/PhasCons_MostConserv_Mammals'; #path
to PhasCons folder containing coordinates of Mammals PhasCons elements sorted
by chromosome and position
        $suffix="MMPPhC";
    }
    elsif($ARGV[1]=~/ver|Ver|VER/){
        $PhasCons_folder='PhasCons/PhasCons_MostConserv_Vertebrates';
#path to PhasCons folder containing coordinates of Vertebrate PhasCons
elements $ARGV[1]=~/m|M/sorted by chromosome and position
        $suffix="VBPhC";
    }
    elsif($ARGV[1]=~/fus|Fus|FUS/){
        $PhasCons_folder='PhasCons/Fused_PhCons/PhasCons_MostConserv_Fused';
#path to PhasCons folder containing coordinates of Fused Vertebrate+Mammals
PhasCons elements sorted by chromosome and position
        $suffix="FsPhC";
    }
    else{die "You must enter PhasCons elements source (Mammals,
Vertebrates,Fused)\n";}
}

else{die "Require chromosome number (example: chr1) and species group (M or V,
for mammals or vertebrates) as arguments\n";}
open (DATA1,
"Chr_processed/Intermediate_files_D/chr".$Chr_name."_Processed.csv") or die
"Cannot open chr".$Chr_name."_Processed.csv file\n";
while(<DATA1>){unless (/name2/){$_=~s/"//g;chomp $_;@tmp=split "\t",$_;push
@Gene_name,$tmp[12];push @Gene_ini,$tmp[13];push @Gene_end,$tmp[14];}}
close(DATA1);
#for($i=0;$i<scalar(@Gene_ini);$i++){print
$Gene_name[$i],"\t",$Gene_ini[$i],"\t",$Gene_end[$i],"\n";}

```

```

open (RES1,
">Chr_processed/Intermediate_files_D/chr".$Chr_name."_Processed_reg.txt") or
die "unable to write processed_reg file\n";
open (RES2,
">Chr_processed/Intermediate_files_D/chr".$Chr_name."_CNS_".$sufix.".txt") or
die "unable to write CNS file\n";
open (RES3,
">Chr_processed/Intermediate_files_D/chr".$Chr_name."_CNS_".$sufix."_filtered.
txt") or die "unable to write CNS_filtered file\n";
open (DATA1,
"Chr_processed/Intermediate_files_D/chr".$Chr_name."_Processed.csv") or die
"Cannot open chr".$Chr_name."_Processed.csv file\n";
while(<DATA1>){
    if (/name2/) {next;};
    chomp $_;$_=~s/"//g;
    print RES1 $_,"\t";
    @tmp=split "\t",$_;
#    print $tmp[12],"\n";
    my @Exon_i=split ",",$tmp[9];my @Exon_e=split ",",$tmp[10];my
@Exon_t=split ",",$tmp[11];
    my ($aa,$bb,$j,@Regi,@Rege);$j=0;@Regi=();@Rege=();
    for ($i=0;$i<scalar(@Exon_t);$i++){
        if ($i==0){
            if
($Exon_t[$i]=~/UTR/){$Regi[$j]=$Exon_i[$i];$Rege[$j]=$Exon_e[$i];}else{$Regi[$
j]=$Exon_i[$i]-1;$Rege[$j]=$Exon_i[$i]-1;}
            $j++;
        }
        elsif($Exon_t[$i]=~/UTR/){
            if ($Exon_e[$i-1]==$Rege[$j-1]){$Rege[$j-1]=$Exon_e[$i];}
            else{
                if (($Exon_i[$i]-$Exon_e[$i-
1])>2){$Regi[$j]=$Exon_e[$i-1]+1;$Rege[$j]=$Exon_e[$i];$j++;}

                else{$Regi[$j]=$Exon_i[$i];$Rege[$j]=$Exon_e[$i];$j++;}
            }
        }
        elsif($Exon_t[$i]=~/CDS/){
            if (($Exon_i[$i]-$Exon_e[$i-1])>2){
                if ($Exon_e[$i-1]==$Rege[$j-1]){$Rege[$j-
1]=$Exon_i[$i]-1;}
                else{$Regi[$j]=$Exon_e[$i-
1]+1;$Rege[$j]=$Exon_i[$i]-1;$j++;}
            }
        }
    }
    # Adds upstream and downstream regions while avoiding overlapping
flanking genes
    if ($tmp[3]=~/\+\/){
        $Regi[0]=&LIMIT_UP($tmp[12],($Regi[0]-
$Up),$Rege[0],\@Gene_name,\@Gene_ini,\@Gene_end);
        $Rege[-1]=&LIMIT_DOWN($tmp[12],$Rege[-1],($Rege[-
1]+$Down),\@Gene_name,\@Gene_ini,\@Gene_end);
    }
    else{
        $Regi[0]=&LIMIT_UP($tmp[12],($Regi[0]-
$Down),$Rege[0],\@Gene_name,\@Gene_ini,\@Gene_end);
        $Rege[-1]=&LIMIT_DOWN($tmp[12],$Rege[-1],($Rege[-
1]+$Up),\@Gene_name,\@Gene_ini,\@Gene_end);
    }
    #Print Non-Coding Regions Boundaries to Gene Table
    for ($i=0;$i<scalar(@Regi);$i++){print RES1 $Regi[$i]," ";}
    print RES1 "\t";
    for ($i=0;$i<scalar(@Rege);$i++){print RES1 $Rege[$i]," ";}
    print RES1 "\n";
    #Identify CNS within Non-Coding Regions
    my @PhCons_ini=();my @PhCons_end=();

```

```

        &FIND_CONS_ELEM($PhasCons_folder,$Chr_name,$#Regi,\@Regi,\@Rege,\@PhCons
_ini,\@PhCons_end);
        for ($i=0;$i<scalar(@PhCons_ini);$i++){print RES2
$tmp[2],"\\t",$PhCons_ini[$i],"\\t",$PhCons_end[$i],"\\n";}
        my @tmp1=(); my @tmp2=();
        &FUSE_CONS_ELEM($min_coverage,$min_l,\@Regi,\@Rege,\@PhCons_ini,\@PhCons
_end,\@tmp1,\@tmp2);
        for ($i=0;$i<scalar(@tmp1);$i++){print RES3
$tmp[2],"\\t",$tmp1[$i],"\\t",$tmp2[$i],"\\n";}

}
close(DATA1);
close(RES1);
close(RES2);
close(RES3);
exit;

sub LIMIT_UP{
    my ($nn,$posi,$pose,$Rn,$Ri,$Re)=@_;
    my $i;
    for ($i=0;$i<scalar(@$Ri);$i++){
        if ($nn eq $$Rn[$i]){next;}
        if ($$Ri[$i]>$posi and $pose>$$Ri[$i]){$posi=$$Ri[$i];}
        if ($$Re[$i]>$posi and $pose>$$Re[$i]){$posi=$$Re[$i];}
    }
    return $posi;
}

sub LIMIT_DOWN{
    my ($nn,$posi,$pose,$Rn,$Ri,$Re)=@_;
    my $i;
    for ($i=0;$i<scalar(@$Ri);$i++){
        if ($nn eq $$Rn[$i]){next;}
        if ($$Ri[$i]>$posi and $pose>$$Ri[$i]){$pose=$$Ri[$i];}
        if ($$Re[$i]>$posi and $pose>$$Re[$i]){$pose=$$Re[$i];}
    }
    return $pose;
}

sub FIND_CONS_ELEM{
    my ($folder,$chr,$ulti,$a_ini,$a_end,$ini_dir,$end_dir)=@_;
    $chr=~s/c/C/;
    $folder=$folder."/chr".$chr;
    my @contador;
    unless(open(DATA2,$folder)){die "file $folder not found";}
    while (<DATA2>){
        chomp $_;$_ =~s/"//g;my @temporal=split"\\t",$_;
        if ($temporal[1]<$a_ini[0]){next;}
        elsif($temporal[0]>$a_end[$ulti]){last;}
        else{
            for ($i=0;$i<scalar(@$a_ini);$i++){
                if ($temporal[0]>=$a_ini[$i] and $temporal[1]<=$a_end[$i]){push
@$ini_dir,$temporal[0];push @$end_dir,$temporal[1];}
                elsif($temporal[0]<$a_ini[$i] and $temporal[1]<=$a_end[$i] and
$temporal[1]>$a_ini[$i]){push @$ini_dir,$a_ini[$i];push
@$end_dir,$temporal[1];}
                elsif($temporal[0]>=$a_ini[$i] and $temporal[0]<$a_end[$i] and
$temporal[1]>$a_end[$i]){push @$ini_dir,$temporal[0];push
@$end_dir,$a_end[$i];}
                elsif($temporal[0]<$a_ini[$i] and $temporal[1]>$a_end[$i]){push
@$ini_dir,$a_ini[$i];push @$end_dir,$a_end[$i];}
            }
        }
    }
    close(DATA2);
}

sub FUSE_CONS_ELEM{

```

## Anexol

```
my ($mc,$ml,$a_ini,$a_end,$ini_dir,$end_dir,$fi,$fe)=@_;
for ($i=0;$i<scalar(@$a_ini);$i++){
    my $ti=0; my $te=0; my $tc=0;
    for ($j=0;$j<scalar(@$ini_dir);$j++){
        unless ($$ini_dir[$j]>=$$a_ini[$i] and
        $$end_dir[$j]<=$$a_end[$i]){next;}
        if ($ti==0 and
        $te==0){$ti=$$ini_dir[$j];$te=$$end_dir[$j];$tc=$te-$ti;}
        else{
            if (($tc+($$end_dir[$j]-
        $$ini_dir[$j]))/($$end_dir[$j]-$ti)>=$mc){
                $tc=$tc+($$end_dir[$j]-
        $$ini_dir[$j]);$te=$$end_dir[$j];
            }
            else{
                if (($te-$ti)>$ml){push @$fi,$ti;push
        @$fe,$te;}
                $ti=$$ini_dir[$j];$te=$$end_dir[$j];$tc=$te-
        $ti;
            }
        }
    }
    if (($te-$ti)>$ml){push @$fi,$ti;push @$fe,$te;}
}
}
```

## NoDupl stored D.pl

```
#!/usr/bin/perl -w
# Eliminate duplicates from chr#_CNS_VBPhC_filtered.txt list generated by
Filter_strategyC.pl

my $chr=$ARGV[0];
    open (DATA,
"Chr_processed/Intermediate_files_D/chr".$chr."_CNS_FsPhC_filtered.txt");
    open (RES, ">Chr_processed/Intermediate_files_D/Duplicated-
entries_chr".$chr."_CNS_FsPhC_filtered.txt");
    my %PCel;
    while(<DATA>){
        chomp $_;$_ =~s/"//g;my @tmp=split "\t",$_;my
        $temp=$tmp[1].".".$tmp[2];
        if (exists($PCel{$temp})) {print RES
        $_,"\n";}else{$PCel{$temp}=$_;}
    }
    close (RES);
    close (DATA);
    my %Done; my $ref;my $cc;
    open
    (RES,">Chr_processed/Intermediate_files_D/chr".$chr."_CNS_FsPhC_filtered_NoDup
    l_sorted.txt");
    until (scalar(keys(%PCel))==scalar(keys(%Done))){
        my $tmpN=1000000000;
        foreach my $i(keys(%PCel)){
            my @tmp1=($i~/(\d+)/g);
            if (exists($Done{$i})) {next;}else{if
        ($tmp1[0]<$tmpN){$tmpN=$tmp1[0];$ref=$i;}}
        }
        print RES $PCel{$ref},"\n";
        print $PCel{$ref},"\n";
        $Done{$ref}=$cc;$cc++;
    }
    close (RES);

exit
```

**Fragment list D.pl**

```

#!/usr/bin/perl -w
# This script reads region tables generated by
Extract_Potential_regulatory_Regions.pl or Filter_cons_reg_table.pl and
generates several independent files containing up to 999 regions
# This is to facilitate extracting the alignments by the UCSC browser (admits
up to 999 regions)
# June 2009
# Luis del Peso
use strict;
my $chr;
if (@ARGV>0){$chr=$ARGV[0];}else{
    print "Enter chromosome number (chr1-chr22 or chrX or chrY):\n";
    $chr=<STDIN>;
    chomp $chr;
}

my
$file_source="Intermediate_files_D/chr".$chr."_CNS_FsPhC_filtered_NoDupl_sorte
d.txt";
mkdir "./Intermediate_files_D/Folder_chr$chr";
my $cont=1;
my $ll;
my $line;

unless (open (DATA, $file_source)){die "File $file_source not found";}
open (RES, ">Chr_processed/Intermediate_files_D/Folder_chr$chr/$cont.txt");
while ($line=<DATA>){
    $ll++;
    if ($ll<998){print RES $line;}
    else{
#        print $cont,"\n";
        print RES $line;
        close (RES);
        $cont++;
        $ll=0;
        open (RES,
">Chr_processed/Intermediate_files_D/Folder_chr$chr/$cont.txt");
    }
}
close (RES);
close (DATA);
exit;

```

**Find HIF Binding sites.pl**

```

#!/usr/bin/perl -w
# This program search Alignments generated by the UCSC genome browser "Tables"
of PhasCons elements contained in potential regulatory regions of genes.
# These regions were identified by the Extract_Potential_regulatory_Regions.pl
program.
# Identifies core regions of a given TF in Homo sapiens sequence and test for
its conservation in the remaining species.
# A site is considered conserved when it is present in the species indicated
in line 190.
# For each conserved site the following parameters are saved: chromosome,
alignment start position, site position within alignment (absolute site
coordinates adding these two), number of species containing the site, sequence
of the site and score of this sequence according to the TF PSSM
# July 2009
# Luis del Peso

```

## Anexol

#### TO USE THIS PROGRAM YOU MUST INDICATE FILE(S) CONTAINING ALIGNMENTS AT LINE 50 AND FILE CONTAINING GENE TABLE AT LINE 76

```
use strict;

my $file_sufix;
my $file;
my $line;
my @temp;
my @tmp;
my @tmpl;
my %Alignm;
my $Hs_seq;
my @Site_pos; #for each potential binding site this variable contains the
coordinates for the first base of the alignment where it is located: The
absolute position for the first base of the core TF binding sequence is
obtained by the sum of this variable and @Coordin
my @Chr; #for each potential binding site this variable contains the
chromosome where it is located
my @Coordin; #for each potential binding site this variable contains the
coordinates for the first base of the alignment where it is located: The
absolute position for the first base of the core TF binding sequence is
obtained by the sum of this variable and @Site_pos
my @N_cons; #for each potential binding site this variable contains the number
of sequences where it is conserved
my @Score; #for each potential binding site this variable contains the score
(according to PSSM) for the binding site
my @TF_seq; #for each potential binding site this variable contains binding
sequence the "N" are added when the core of the binding site is close to the
ends of the alignment
my $Hs_Chr;
my $Hs_pos;
my $Hs_pos_i;
my $tmpLong;
my $ll;
my $l;
my ($j,$i);
my $min_size=10; #minimal alignment length
my $PSSM_Core_L=5;
my @PSSM; #Contains Position Specific Score Matrix
my $matrix_file='PSSMlogodd025-8to+17.txt';
my @NM_ref; #contains gene ID
my @Gene_name; #contains gene name
my $upstream=5000; #number of bases upstream transcription start site
my $downstream=2000; #number of bases downstream last exon
my $Extra_bp=25; #number of bases to extend alignment during processing with
Filter_ConsReg_v3.pl program
my @Region; #Contains the region where the core TF binding site is located.
my %Visto;
my $chrN; #Contains chr number to process
if (@ARGV){$chrN=$ARGV[0];}else{print "Enter chr number (1-22 or X or
Y)\n";$chrN=<STDIN>;chomp $chrN;}

leePSSM(\@PSSM,$matrix_file); #Subrutina Lee PSSM y la almacena en @PSSM
#print scalar(@PSSM)," \n";
#print "##### MATRIZ #####\n";
#for ($j=1;$j< scalar (@PSSM);$j++)
#{
#print
$PSSM[$j]{A}," \t",$PSSM[$j]{C}," \t",$PSSM[$j]{G}," \t",$PSSM[$j]{T}," \n";
#}
open (RES, ">CNS_containing_HREs_HIF_Targets_chr".$chrN.".txt") or die "unable
to write on RES\n";
chdir ("Alignm_PhaseCons_D/alineam_Chr".$chrN."/");
my @af=glob("alin*");
chdir ("../..");
foreach $file_sufix(@af){
    #unless($file_sufix=~/_1\.txt/){next;}
}
```



```

$file="Alignm_PhaseCons_D/alineam_Chr".$chrN."/". $file_sufix;
print $file,"\\n";
unless(open(DATA, $file)){die "file $file not found";}
while ($line=<DATA>){
    my @temp=split (" ", $line);
    if ($line=~/^s/ and $line=~chr/ and !exists($Alignm{hg18}))){#Saves
aligment in has %Alignm
        $temp[1]=~/([0-9a-zA-Z]+)\\.chr([0-9a-zA-Z]+)/;
        $Alignm{$1}=uc($temp[-1]);
        if ($1 eq "hg18"){ $Hs_Chr=$2; $Hs_pos=$temp[-5]; $l1=$temp[-
4]; $Hs_pos_i=$temp[-5];}
    }
    elsif ($line=~/^s/ and $line=~chr/ and !($line=~hg18/))){#Saves
aligment in has %Alignm
        $temp[1]=~/([0-9a-zA-Z]+)\\.chr([0-9a-zA-Z]+)/;
        if ($Alignm{$1}){ $Alignm{$1}=$Alignm{$1}.uc($temp[-
1]);}else{ $Alignm{$1}=uc($temp[-1]);}
    }
    elsif($line=~/^e/ and $line=~chr/ and $line=~ /\\d+\\sI|\\d+\\sM/ and
exists($Alignm{hg18}))){#Includes string of "-" when insertion is found
        $temp[1]=~/([0-9a-zA-Z]+)\\.chr([0-9a-zA-Z]+)/;
        if ($Alignm{$1}){ $Alignm{$1}=$Alignm{$1}."-
"x$tmpLong;}else{ $Alignm{$1}="-"x(length($Alignm{hg18}));}
        #if($Hs_pos_i>=87921780 and $Hs_pos_i<87921996){print
$1,"\\t",$Hs_pos,"\\t",$Hs_pos_i,"\\t",$tmpLong,"\\t", (length($Alignm{$1})), "\\n", $
Alignm{hg18}, "\\n";}
    }
    elsif($line=~/^s\\s+hg18/ and $Alignm{hg18}){#End of Aligment
        #print "----->",$line;
        @temp=split (" ", $line);
        if ($temp[-5]==($Hs_pos+$l1)){
            $Alignm{hg18}=$Alignm{hg18}.uc($temp[-1]);
            $Hs_pos=$temp[-5]; $l1=$temp[-4]; $tmpLong=length($temp[-1]);
            next;
        }
        else{
            $Hs_seq=$Alignm{hg18};
            $Hs_seq=~s/-//g;
            #print RESULTS $Hs_Chr,"\\t",$Hs_pos,"\\t",$Hs_seq,"\\n";
            #print $Hs_Chr,"\\t",$Hs_pos,"\\t",$Hs_seq,"\\n";
            #print "Aligment found at $Hs_pos\\t";
            if (length ($Hs_seq)>$min_size and ($Hs_seq=~/[AG]CGTG/ or
$Hs_seq=~CACG[TC]/)){
                # print $Hs_seq,"\\n";
                # print "CONTAIN potential HRE\\n";
                print RES $Hs_seq,"\\n";
            }
        }
    }
    $Hs_seq=$Alignm{hg18};
    $Hs_seq=~s/-//g;
    if (length ($Hs_seq)>$min_size and ($Hs_seq=~/[AG]CGTG/ or
$Hs_seq=~CACG[TC]/)){
        # print $Hs_seq,"\\n";
        # print "CONTAIN potential HRE\\n";
        print RES $Hs_seq,"\\n";
    }
}
TEST_CONS($Hs_seq,$Hs_Chr,$Hs_pos_i,\\%Alignm,\\@Chr,\\@Coordin,\\@Site_pos,\\@N_co
ns,\\@Score,\\@TF_seq,$PSSM_Core_L,\\@PSSM,\\%Visto);

```

```

    }
    %Alignm=();
    close (DATA);
}
close (RES);
#print
scalar (@Chr), "\t", scalar (@Coordin), "\t", scalar (@Site_pos), "\t", scalar (@N_cons)
, "\t", scalar (@Score), "\t", scalar (@TF_seq), "\n";
open (RESULTS, ">TEST_chr".$chrN.".csv") or die "unable to write TEST\n";
for ($i=0; $i<scalar (@Chr); $i++) {print RESULTS
$Chr[$i], "\t", $Coordin[$i], "\t", $Site_pos[$i], "\t", $N_cons[$i], "\t", $Score[$i]
, "\t", $TF_seq[$i], "\n";}
close (RESULTS);
unless (open (DATA,
("Intermediate_files_D/chr".$chrN."_Processed_reg.txt"))) {die "unable to open
./Chr_processed/Intermediate_files/chr".$chrN."_Processed_reg.txt\n";};
open (RESULTS, ">Conserved_elements_HIFTargets_chr".$chrN.".csv") or die
"unable to write Conserved_elements\n";
print RESULTS
"chr\tGene\tcoordinates\tposition\tNspecies\tsequence\tscore\tmax\n";
while (<DATA>){
    if ($_ =~ /txStart/) {next;}
    chomp $_; $_ =~ s/"//g; @tmp=split "\t", $_;
    chop ($tmp[16]); chop ($tmp[17]); my @ll=split ",", $tmp[16]; my @rr=split
    ",", $tmp[17];
    #if ($tmp[12] eq "GYS1") {print
$tmp[16], "\t", $ll[0], "\n", $tmp[17], "\t", $rr[-1], "\n"; exit;}
    my
($cc, $pp, $nc, $sq, $sc, $mm) = &MAP_to_LOCUS ($tmp[2], $tmp[12], $tmp[13], $tmp[14], $ll
[0], $rr[-1], \@Chr, \@Coordin, \@Site_pos, \@N_cons, \@Score, \@TF_seq);
    print RESULTS
$tmp[2], "\t", $tmp[12], "\t", $cc, "\t", $pp, "\t", $nc, "\t", $sq, "\t", $sc, "\t", $mm, "\
n";
}
close (RESULTS);
close (DATA);
exit;

sub TEST_CONS{
    my
($Hs, $H_C, $H_P, $dir, $dir_ch, $dir_co, $dir1, $dir2, $dir3, $dir4, $PL, $PSSM_ref, $vv)
=@_;
    my $xx;
    my $yy=-1;
    my $ii;
    my $ll;
    my $n;
    my $cc;
    my $jj=0;
    my $Pos_ini;
    my $Pos_end;
    my $esp;
    my %Sp_cons=();
    my $dif;
    my $HsN;
    my @Spp=keys (%$dir);
#    print
$H_P, "\t", substr ($$dir{hg18}, 0, 5), "\t", length ($$dir{hg18}), "\t", substr ($$dir{h
g18}, -5, 5), "\n";
    #for ($ll=0; $ll<scalar (@Spp); $ll++) {print $Spp[$ll], "\n";} print
"*****\n";
    my @sites=($Hs =~ / [AG]CGTG | CACG [TC] /g);
    #for ($ll=0; $ll<scalar (@sites); $ll++) {print
$H_C, "\t", $H_P, "\t", $sites[$ll], "\n";} print
"*****\n";
    foreach $ii (@sites) { ### Evaluate each core sequence in Human
        $xx=index ($Hs, $ii, $yy+1);
        $yy=$xx;

```

```

$cc=0;
#   print $ii,"\t", $xx, "\n";
$Pos_ini=0;
$Pos_end=0;
for ($n=0;$n<length($$dir{hg18});$n++){
  #print $n, "\t", substr($$dir{hg18}, $n, 1), "\t", $cc, "\t";
  if ($cc==$xx){$Pos_ini=$n;}
  if ($cc==($xx+$PL)){ $Pos_end=$n; last;}
  unless (substr($$dir{hg18}, $n, 1) eq "-"){ $cc++;}
  #print $cc, "\t", $Pos_ini, "\t", $Pos_end, "\n";
}
if ($cc==($xx+$PL)){ $Pos_end=$n;}
#   print
$H_C, "\t", $H_P, "\t", $Pos_ini, "\t", $Pos_end, "\t", $ii, "\t", $xx, "\t", substr($$dir
{hg18}, $Pos_ini, $Pos_end-$Pos_ini), "\n";
  $dif=$Pos_end-$Pos_ini;
#   print
$H_C, "\t", $H_P, "\t", $Pos_ini, "\t", $Pos_end, "\t", $ii, "\t", $xx, "\t", substr($$dir
{hg18}, $Pos_ini, $dif), "\n";
  %Sp_cons=();
  foreach $esp (@Spp){
    #print $esp, "\n";
    if (length($$dir{$esp}) < ($Pos_ini+$dif)) {next;}
    else{
#       print $esp, "\t", substr($$dir{$esp}, $Pos_ini, $dif), "\n";
        if (substr($$dir{hg18}, $Pos_ini, $dif) eq
substr($$dir{$esp}, $Pos_ini, $dif)){
          $jj++;
          $Sp_cons{$esp}=substr($$dir{$esp}, $Pos_ini, $dif);
#         print "****
", $esp, "\t", substr($$dir{$esp}, $Pos_ini, $dif), "\n";
        }
      }
    }
  }
  if (exists($Sp_cons{mm8}) and (exists($Sp_cons{rn4}) or
exists($Sp_cons{canFam2}) or exists($Sp_cons{bosTau3}) or
exists($Sp_cons{eriEur1}) or exists($Sp_cons{loxAfr1}) or
exists($Sp_cons{monDom4})) and (exists($Sp_cons{rheMac2}) or
exists($Sp_cons{panTro2}))) {
    unless ($$vv{$H_C.$H_P.$xx}) {
#       print "SITE FOUND\n";
      push @dir1, $xx;
      push @dir2, $jj;
      push @dir_ch, $H_C;
      push @dir_co, $H_P;
      $HsN="NNNNNNNNNNNNNNNNNNNNNNNNNNNNNN".$Hs."NNNNNNNNNNNNNNNNNNNNNNNNNNNNNN";
      if (($ii eq "ACGTG") or ($ii eq
"GCGTG")) {$HsN=substr($HsN, $xx+12, 25);}
      else {$HsN=REVCPL(substr($HsN, $xx+8, 25));}
      #print $HsN, "\n";
      push @dir3, PSSMScore($HsN, $PSSM_ref);
      push @dir4, $HsN;
      #print $$dir1[-1], "\t", $$dir2[-1], "\t", $$dir_ch[-
1], "\t", $$dir_co[-1], "\t", $$dir3[-1], "\t", $$dir4[-1], "\n";
      $$vv{$H_C.$H_P.$xx}="seen";
    }
  }
}
#   print "Sites found so far: ", scalar (@dir_ch), "\n";
}

sub leePSSM {
  my ($ref_PSSM, $file_matrix) = @_;
#Store the Position-specific scoring matrix of frequencies en un array en el
que cada posicion es un elemento del array compuesto de un hash que asigna la
prob de cada nucleotido

```

```

unless (open (ARCHIVO1, $file_matrix)){die "Can't open file $file_matrix
containing PSSM.\n";}
my @matrix_data=<ARCHIVO1>;
close (ARCHIVO1);
#print scalar (@temp),"\n";
#foreach (@temp)
#{
#print $temp[$i];
#$i++;
#}
my @temp_2;
my $i=0;
for ($i=0;$i<scalar(@matrix_data);$i++)
{
    chomp $matrix_data[$i];
    @temp_2= split /\t/, $matrix_data[$i];
    #print $temp[$i],"\n","last ", $temp2[3],"\n";
    push @$ref_PSSM, {A=>$temp_2[0],
C=>$temp_2[1],G=>$temp_2[2],T=>$temp_2[3]}
}
}

sub REVCPL{
my ($ssqq) = @_ ;
#print $tmp1,"\n";
my @tmp3=split ('',$ssqq);
my $l=0;
my $sc="";
for ($l=0;$l<scalar(@tmp3);$l++){
    if ($tmp3[$l] eq "A"){ $sc="T".$sc;}
    elsif ($tmp3[$l] eq "T"){ $sc="A".$sc;}
    elsif ($tmp3[$l] eq "C"){ $sc="G".$sc;}
    elsif ($tmp3[$l] eq "G"){ $sc="C".$sc;}
    else{ $sc=$tmp3[$l].$sc;}
}
#print $tmp3[$l],"->",$PSSM[$l]{ $tmp3[$l] },"\n";
#print "\n",$sc,"\n";
return $sc;
}

sub PSSMScore {
my ($Secuencia,$reff) = @_ ;
#print $tmp1,"\n";
my @tmp3=split ('',$Secuencia);
my $l=0;
my $sc=0;
for ($l=0;$l<scalar(@tmp3);$l++){
    if (exists($$reff[$l]{ $tmp3[$l] })){
        $sc=$sc+$$reff[$l]{ $tmp3[$l] };
        #print $tmp3[$l],"->",$$reff[$l]{ $tmp3[$l] },"\n";
    }
}
#print "\n",$sc,"\n";
return ($sc);
}

sub SORT_Sites{
my ($So,$Nc,$Sc)=@_ ;
my $max=-100;
my $Pos_max=0;
my $t;
my $r=0;
my %Ss;
while ($r<scalar(@$Sc)){
    $max=-100;
    for ($t=0;$t<scalar(@$Sc);$t++){
#
        print
$r,"\t",$Ss[$t],"\t",$Nc[$t],"\t",$max,"\t",$Pos_max,"\t";

```

```

        if (($Sc[$t]+$Nc[$t])>=$max and
!exists($Ss{$t})){$max=$Sc[$t];$Pos_max=$t;}
#       print $max,"\t",$Pos_max,"\n";
    }
    push @$So, $Pos_max;
    $Ss{$Pos_max}=1;
    $r++;
}
}

sub MAP_to_LOCUS{
    my ($cr,$gs,$li,$le,$ri,$re,$dchr,$dco,$dpos,$dNcons,$dscr,$dseq)=@_ ;
    my $k; my $hre=0;my $rc;my $rp;my $rnc;my $rsq;my $rscr;my $max=-100;my
%eliminate;
    for ($k=0;$k<scalar(@$dchr);$k++){
        if (($cr eq ("chr".$dchr[$k])) and ($ri<($dco[$k]+$dpos[$k]))
and ($re>($dco[$k]+$dpos[$k]))){
            $hre=1;
            $eliminate{$k}=1;
            if ($rc){$rc=$rc."|".$dco[$k];}else{$rc=$dco[$k];}
            if ($rp){$rp=$rp."|".$dpos[$k];}else{$rp=$dpos[$k];}
            if
($rnc){$rnc=$rnc."|".$dNcons[$k];}else{$rnc=$dNcons[$k];}
            if ($rsq){$rsq=$rsq."|".$dseq[$k];}else{$rsq=$dseq[$k];}
            if
($rscr){$rscr=$rscr."|".$dscr[$k];}else{$rscr=$dscr[$k];}
            if ($dscr[$k]>$max){$max=$dscr[$k];}
        }
    }
    #for ($k=0;$k<scalar(@$dchr);$k++){if ($eliminate{$k}){splice
@dchr,$k,1;splice @dco,$k,1;splice @dpos,$k,1;splice @dNcons,$k,1;splice
@dscr,$k,1;splice @dseq,$k,1;}}
    if ($hre>0){return ($rc,$rp,$rnc,$rsq,$rscr,$max);}else{return
("NA","NA","NA","NA","NA","NA");}
}

```

### Combine all data.pl

```

#!/usr/bin/perl -w
use strict;
my $i;my $j=0;my @tmp;
open (RES, ">Combined_data_curated.csv") or die "Unable to create file\n";
print RES
"Chr\tGene\tcoordinates\tposition\tNspecies\tsequence\tscore\tmax\tGene_Symbol
\tp_value\tln\tMean\tSD\tpvalue_adjusted\n";
foreach my $cc (1..22,"X","Y"){
    if ($cc==19){open (DATA1,
"chr".$cc."_curated/Conserved_elements_HIFTargets_chr".$cc."_curated.csv") or
die "file not found\n";}else{open (DATA1,
"chr".$cc."/Conserved_elements_HIFTargets_chr".$cc.".csv") or die "file not
found\n";}
    while(<DATA1>){
        $j++;
        if ($_~/Nspecies/){next;}
        chomp $_;$_ =~ s/"//g;@tmp=split "\t",$_;
        print $j,"\t",$tmp[1],"\t";
        print RES $_,"\t",&GEO_data($tmp[1]),"\n";
    }
    close (DATA1);
}
close (RES);
exit;

sub GEO_data{
    my ($Gsym)=@_ ;
    my $ff;
    print $Gsym,"\n";
}

```

## Anexol

```
open (DATA2, "/home/luis/Documentos/BIOINFO/My Programs/Perl
programs/GEO/FDR/pvalues_standarized_extended/All_pvalues_standarized_extended
_adjusted.csv") or die "file not found\n";
while($ff=<DATA2>){
    chomp $ff;$ff=~s/"//g;my @tempo=split "\t",$ff;
    if ($tempo[0] eq $Gsym){close (DATA2); return $ff;}
}
close(DATA2);
return ("NOT FOUND IN GEO DATA SETS");
}
```

## Get human seq v4.pl

```
#!/usr/bin/perl -w
# Gets the whole human sequence containing identified potential HREs
# August 2009
# Luis del Peso

use LWP::UserAgent;
$ua=LWP::UserAgent->new;
$ua->agent("MyApp");
my %GS;
open (DATA, "missing_genes.csv") or die "file not found\n";
while (<DATA>){
    chomp $_;$_=~s/"//g;my @tmp=split "\t", $_;$tmp[0]=~/[A-Z0-9-]//g;
    if ($tmp[2] eq "NO" and $tmp[0]=~/[A-Z]/){$GS{$tmp[0]}=$tmp[0];}
}
close(DATA);
open (DATA, "Combined_data_curated.csv") or die "file not found\n";
open (RES, ">Combined_data_curated_HRE_Seq_missing.csv") or die "unable to
open file\n";
my $j;my $ok=0;my $ok2=0;
while (<DATA>){
    #print ">>>>>>$_ ";
    chomp $_;$_=~s/"//g;my @tmp=split "\t", $_;
    if ($_~/coordinates/){print RES $_,"\tWhole_seq","\n";next;}
    unless (exists ($GS{$tmp[1]})) {next;}
    $tmp[0]=~/chr//;print "#####\t",$tmp[1],"\n";
#    if ($tmp[1] eq "NUMB") {$ok2=1;}
#    unless ($ok2>0) {next;}
#    my ($k,$li)=&Ya_leido($tmp[1]);
#    if ($k>0) {print RES $li;next;}
#    unless ($ok>0) {next;}
    my @tmp2=($tmp[2]=~/(\d+)/g);my @tmp3=($tmp[3]=~/(\d+)/g);my
@tmp5=($tmp[5]=~/([ACGTN]){25})/g);
    if ($tmp[2]=~/NA/){print RES $_,"\tNA\n";print "NA--->NEXT\n";next;}
    my $Wseq33="";
    for ($j=0;$j<scalar(@tmp2);$j++){
        unless ($tmp5[$j]=~/N/){if($Wseq33 eq
""){$Wseq33=$tmp5[$j];}else{$Wseq33=$Wseq33."|".$tmp5[$j];}next;}
        my $ini=$tmp2[$j]+$tmp3[$j]-26;my
$end=$tmp2[$j]+$tmp3[$j]+22;
        #my $dir='http://genome.ucsc.edu/cgi-
bin/hgc?hgsid=125551557&g=htcGetDna2&table=&i=mixed&o=54175765&l=54175765&r=54
175790&getDnaPos=chr'.$tmp[0].'%3A'.$ini.'-
'.$end.'&db=hg18&hgSeq.cdsExon=1&hgSeq.padding5=&hgSeq.padding3=&hgSeq.casing=
upper&boolshad.hgSeq.maskRepeats=1&hgSeq.repMasking=lower&boolshad.hgSeq.revCo
mp=1&submit=get+DNA';
        #my $dir='http://genome.ucsc.edu/cgi-
bin/hgc?hgsid=137292352&g=htcGetDna2&table=&i=mixed&o=6512929&l=6512929&r=6512
973&getDnaPos=chr'.$tmp[0].'%3A'.$ini.'-
'.$end.'&db=hg18&hgSeq.cdsExon=1&hgSeq.padding5=&hgSeq.padding3=&hgSeq.casing=
upper&hgSeq.maskRepeats=on&boolshad.hgSeq.maskRepeats=0&hgSeq.repMasking=lower
&boolshad.hgSeq.revComp=0&submit=get+DNA';
```

```

my $dir='http://genome.ucsc.edu/cgi-
bin/hgc?hgsid=139432162&g=htcGetDna2&table=&i=mixed&o='.( $ini-1).'&l='.( $ini-
1).'&r='.$end.'&getDnaPos=chr'.$tmp[0].'%3A'.$ini.'-
'.$end.'&db=hg18&hgSeq.cdsExon=1&hgSeq.padding5=0&hgSeq.padding3=0&hgSeq.casin
g=upper&boolshad.hgSeq.maskRepeats=0&hgSeq.repMasking=lower&boolshad.hgSeq.rev
Comp=0&submit=get+DNA';
my $req=HTTP::Request->new(GET=>$dir);
my $res=$ua->request($req);
if ($res->is_success){
    #print $res->content;
    my $resdata=$res->content;
    print $resdata;
    @resdata=split "\n", $resdata;
    chomp $resdata[2];chomp $resdata[3];
    print
"*****",$resdata[2],"\n*****",$resdata[3],"\n";
    #print RES
$_,"\t",$resdata[2],"\t",$resdata[3],"\n";
    if($Wseq33 eq
""){ $Wseq33=$resdata[3];}else{ $Wseq33=$Wseq33."|".$resdata[3];}
    }
    else{print $res->status_line,"\n";}
    sleep (4);
}
print RES $_,"\t",$Wseq33,"\n";
}
close (RES);
close (DATA);
exit;
sub Ya_leido{
my ($Symb)=@_;
open (DATA1, "Combined_data_Dv2_HRE_Seq.csv") or die "file not found\n";
while(my $line=<DATA1>){
    chomp $line;$line=~s/"/"/g;my @temp=split"\t",$line;
    if ($temp[1] eq $Symb){close (DATA1);return(1,$line);}
    #sleep 2; print "FOUND IN SUB\n";
}
close (DATA1);
return(0,"");
}
#http://genome.ucsc.edu/cgi-
bin/hgc?hgsid=137292237&g=htcGetDna2&table=&i=mixed&o=6512939&l=6512939&r=6512
973&getDnaPos=chr12%3A6%2C512%2C940-
6%2C512%2C973&db=hg18&hgSeq.cdsExon=1&hgSeq.padding5=&hgSeq.padding3=&hgSeq.ca
sing=upper&hgSeq.maskRepeats=on&boolshad.hgSeq.maskRepeats=0&hgSeq.repMasking=
lower&boolshad.hgSeq.revComp=0&submit=get+DNA
#http://genome.ucsc.edu/cgi-
bin/hgc?hgsid=137292352&g=htcGetDna2&table=&i=mixed&o=6512929&l=6512929&r=6512
973&getDnaPos=chr12%3A6%2C512%2C930-
6%2C512%2C973&db=hg18&hgSeq.cdsExon=1&hgSeq.padding5=&hgSeq.padding3=&hgSeq.ca
sing=upper&hgSeq.maskRepeats=on&boolshad.hgSeq.maskRepeats=0&hgSeq.repMasking=
lower&boolshad.hgSeq.revComp=0&submit=get+DNA
#http://genome.ucsc.edu/cgi-
bin/hgc?hgsid=125551557&g=htcGetDna2&table=&i=mixed&o=54175765&l=54175765&r=54
175790&getDnaPos=chr'.$tmp[0].'%3A'.$ini.'-
'.$end.'&db=hg18&hgSeq.cdsExon=1&hgSeq.padding5=&hgSeq.padding3=&hgSeq.casing=
upper&boolshad.hgSeq.maskRepeats=1&hgSeq.repMasking=lower&boolshad.hgSeq.revCo
mp=1&submit=get+DNA'

#http://genome.ucsc.edu/cgi-
bin/hgc?hgsid=137292352&g=htcGetDna2&table=&i=mixed&o=6512929&l=6512929&r=6512
973 &getDnaPos=chr'.$tmp[0].'%3A'.$ini.'-'. $end.'
&db=hg18&hgSeq.cdsExon=1&hgSeq.padding5=&hgSeq.padding3=&hgSeq.casing=upper&hg
Seq.maskRepeats=on&boolshad.hgSeq.maskRepeats=0&hgSeq.repMasking=lower&boolsha
d.hgSeq.revComp=0&submit=get+DNA
#http://genome.ucsc.edu/cgi-
bin/hgc?hgsid=139432162&g=htcGetDna2&table=&i=mixed&o=124965479&l=124965479&r=

```

```

124969889&getDnaPos=chr11%3A124%2C965%2C480-
124%2C969%2C889&db=hg18&hgSeq.cdsExon=1&hgSeq.padding5=0&hgSeq.padding3=0&hgSeq.casing=upper&boolshad.hgSeq.maskRepeats=0&hgSeq.repMasking=lower&boolshad.hgSeq.revComp=0&submit=get+DNA
#http://genome.ucsc.edu/cgi-bin/hgc?hgsid=139432162&g=htcGetDna2&table=&i=mixed&o=33463171&l=33463171&r=33463180&getDnaPos=chr14%3A33%2C463%2C172-33%2C463%2C180&db=hg18&hgSeq.cdsExon=1&hgSeq.padding5=0&hgSeq.padding3=0&hgSeq.casing=upper&boolshad.hgSeq.maskRepeats=0&hgSeq.repMasking=lower&boolshad.hgSeq.revComp=0&submit=get+DNA
#http://genome.ucsc.edu/cgi-bin/hgc?hgsid=139432162&g=htcGetDna2&table=&i=mixed&o=54163194&l=54163194&r=54163200&getDnaPos=chr19%3A54%2C163%2C195-54%2C163%2C200&db=hg18&hgSeq.cdsExon=1&hgSeq.padding5=0&hgSeq.padding3=0&hgSeq.casing=upper&boolshad.hgSeq.maskRepeats=0&hgSeq.repMasking=lower&boolshad.hgSeq.revComp=0&submit=get+DNA
#http://genome.ucsc.edu/cgi-bin/hgc?hgsid=139432162&g=htcGetDna2&table=&i=mixed&o='.( $ini-1 ).'&l='.( $ini-1 ).'&r='.$end.'&getDnaPos=chr'.'$tmp[0]'. '%3A'.'$ini.'-'. $end.'&db=hg18&hgSeq.cdsExon=1&hgSeq.padding5=0&hgSeq.padding3=0&hgSeq.casing=upper&boolshad.hgSeq.maskRepeats=0&hgSeq.repMasking=lower&boolshad.hgSeq.revComp=0&submit=get+DN

```

### W Scores PSSM 46seq.pl

```

#!/usr/bin/perl -w
# calculates HRE scores
# August 2009
# Luis del Peso

use strict;
my $i;
my $Pos_i=1;my $Total_L=18;
my $matrix_file='PSSMs_46seq_1_to_16.csv';my @PSSM;
leePSSM(\@PSSM,$matrix_file); #Subrutina Lee PSSM y la almacena en @PSSM
open (DATA,"Combined_data_curated_HRE_Seq_ok2.csv") or die "DATA not found\n";
open (RES,">Combined_data_curated_HRE_Seq_ok2_wScores_46seq.csv") or die
"Unable to write file\n";
while(<DATA>){
    chomp $_;$_=~s/"//g;my @tmp=split"\t",$_;
    if ($_~/coordinates/){print RES
$_,"\tTrim_Seq\twScore\tMax_wScore\n";next;}
    $tmp[14]=uc($tmp[14]);$tmp[5]=~s/N//g;
    if ($tmp[8]=~/NOT/){print RES $_,"\t"x6,"NA\tNA\tNA\n";next;}
    if ($tmp[14] eq "NA"){print RES $_,"\tNA\tNA\tNA\n";next;}
    my @tmp1=($tmp[5]=~/([ACGT]+)/g);my @tmp2=($tmp[14]=~/([ACGT]+)/g);
    my $wScore="";my $MaxScore=-100;my $Tseq="";
    print $tmp[1],"\n";
    for($i=0;$i<scalar(@tmp2);$i++){
        my $j;my $score;my $seq;
        if
(length($tmp2[$i])<45){$seq=substr($tmp1[$i],7,18);$seq=uc($seq);$score=PSSMScore($seq,\@PSSM);}
        elsif (substr($tmp2[$i],25,7)=~/ACGT/){
            $j=index($tmp2[$i],"ACGT",25);
            if (substr($tmp2[$i],$j-1,6) eq "CACGTG"){
                $seq=substr($tmp2[$i],$j-
$Pos_i,$Total_L);$seq=uc($seq);
                $score=PSSMScore($seq,\@PSSM);
                my $Sc=PSSMScore(uc(REVCPL(substr($tmp2[$i],$j-
$Total_L+$Pos_i+4,$Total_L))),\@PSSM);
                if
($Sc>$score){$score=$Sc;$seq=REVCPL(substr($tmp2[$i],$j-
$Total_L+$Pos_i+4,$Total_L));$seq=uc($seq);}

```



```

    }
    elsif(substr($tmp2[$i],$j,5) eq "ACGTG"){
        $seq=substr($tmp2[$i],$j-
$Pos_i,$Total_L);$seq=uc($seq);
        $score=PSSMScore($seq,\@PSSM);
    }
    else{
        $seq=REVCPL(substr($tmp2[$i],$j-
$Total_L+$Pos_i+4,$Total_L));$seq=uc($seq);
        $score=PSSMScore($seq,\@PSSM);
    }
}
elseif (substr($tmp2[$i],25,7)=~/GCGT/){
    $j=index($tmp2[$i],"GCGT",22);
    $seq=substr($tmp2[$i],$j-$Pos_i,$Total_L);$seq=uc($seq);
    $score=PSSMScore($seq,\@PSSM);
}
elseif (substr($tmp2[$i],25,7)=~/ACGC/){
    $j=index($tmp2[$i],"ACGC",22);
    $seq=REVCPL(substr($tmp2[$i],$j-
$Total_L+$Pos_i+4,$Total_L));$seq=uc($seq);
    $score=PSSMScore($seq,\@PSSM);
}
else{$seq="NOT FOUND";$score=0;}
print "\t",$seq,"\t",$score,"\n";
if ($wScore eq
""){ $wScore=$score;$Tseq=$seq;}else{ $wScore=$wScore."|". $score;$Tseq=$Tseq."|"
.$seq;}
    if ($score>$MaxScore){ $MaxScore=$score;}
}
print RES $_,"\t",$Tseq,"\t",$wScore,"\t",$MaxScore,"\n";
}
close(RES);
close(DATA);
exit;
sub leePSSM {
    my ($ref_PSSM,$file_matrix)=@_;
    #Store the Position-specific scoring matrix of frequencies en un array
en el que cada posicion es un elemento del array compuesto de un hash que
asigna la prob de cada nucleotido
    unless (open (ARCHIVO1, $file_matrix)){die "Can't open file $file_matrix
containing PSSM.\n";}
    my @matrix_data=<ARCHIVO1>;
    close (ARCHIVO1);
    #print scalar (@temp),"\n";
    #foreach (@temp)
    #{
    #print $temp[$i];
    # $i++;
    #}
    my @temp_2;
    my $i=0;
    for ($i=0;$i<scalar(@matrix_data);$i++){
        chomp $matrix_data[$i];
        @temp_2= split /\t/, $matrix_data[$i];
        #print $temp[$i],"\n","last ", $temp2[3],"\n";
        push @$ref_PSSM, {A=>$temp_2[0],
C=>$temp_2[1],G=>$temp_2[2],T=>$temp_2[3]}
    }
}

sub REVCPL{
    my ($ssqq) = @_;
    #print $tmp1,"\n";
    my @tmp3=split ('',$ssqq);
    my $l=0;
    my $sc="";
    for ($l=0;$l<scalar(@tmp3);$l++){

```

```

        if ($tmp3[$1] eq "A"){$sc="T".$sc;}
        elsif ($tmp3[$1] eq "T"){$sc="A".$sc;}
        elsif ($tmp3[$1] eq "C"){$sc="G".$sc;}
        elsif ($tmp3[$1] eq "G"){$sc="C".$sc;}
        else{$sc=$tmp3[$1].$sc;}
    }
    #print $tmp3[$1],"->", $PSSM[$1]{$tmp3[$1]},"\n";
    #print "\n", $sc, "\n";
    return $sc;
}

sub PSSMScore {
    my ($Secuencia,$reff) = @_ ;
    #print $tmp1,"\n";
    my @tmp3=split ('',$Secuencia);
    my $l=0;
    my $sc=0;
    for ($l=0;$l<scalar(@tmp3);$l++){
        if (exists($$reff[$l]{$tmp3[$l]})){
            $sc=$sc+$$reff[$l]{$tmp3[$l]};
            #print $tmp3[$l],"->", $$reff[$l]{$tmp3[$l]},"\n";
        }
    }
    #print "\n", $sc, "\n";
    return ($sc);
}

```

IX. ANEXO II

Anexo II. Matriz de frecuencias generada a partir de 45 secuencias diana de HIF.

GEN	Sp	SECUENCIA																																																							
		-16	-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21																			
ENO1	Hs	G	T	G	C	G	G	A	C	T	C	G	G	A	G	T	A	C	G	T	G	G	A	C	G	G	A	G	A	C	C	C	C	C	G	A	A	C	C	A	A	C	C	C	C												
ENO1	Cf	T	C	C	C	A	C	C	C	C	C	C	C	T	C	G	A	C	G	T	G	A	C	T	C	G	A	C	C	A	A	T	C	C	A	A	T	C	C	A	A	C	C	C	C												
ALDOA	Hs	C	C	C	C	G	C	C	C	C	C	C	C	T	A	A	G	A	C	G	T	G	A	C	T	C	C	A	A	T	C	C	A	A	T	C	C	A	A	T	C	C	C	T	C												
ALDOA	Cf	C	C	C	C	G	G	C	C	C	C	C	G	T	G	T	A	A	C	G	T	G	C	A	G	A	G	C	A	A	G	A	A	C	A	A	G	A	A	G	A	C	A	G	C	G											
EGLN1	Hs	G	C	C	G	G	G	C	C	G	T	G	G	T	G	T	A	A	C	G	T	G	C	A	G	C	G	C	G	C	G	C	C	A	A	G	A	G	A	A	C	A	G	C	G	C											
EGLN1	Mm	G	C	T	G	A	G	T	G	A	C	A	G	T	G	C	T	A	C	G	T	G	C	A	G	C	G	C	G	C	G	C	A	A	G	A	G	A	G	A	A	C	A	G	C	G											
EGLN3	Hs	G	C	T	A	A	T	G	A	C	A	G	T	G	C	T	A	A	C	G	T	G	C	A	G	C	G	C	G	C	T	A	A	G	A	G	A	G	C	T	A	T	C	A	T	C	T										
EGLN3	Gg	C	C	G	C	G	C	A	C	G	C	G	C	C	C	G	C	A	C	G	T	G	C	A	C	A	C	C	C	C	C	A	C	C	A	C	C	C	C	C	C	C	C	A	C	C	G										
BNIP3	Hs	C	C	C	G	C	C	A	C	G	C	G	C	C	C	G	C	A	C	G	T	G	C	A	C	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T	G									
BNIP3	Rn	C	C	C	G	C	C	A	C	G	C	G	C	C	C	G	C	A	C	G	T	G	C	A	C	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T	G									
P4Ha	Hs	G	A	C	T	G	C	A	G	C	C	T	C	G	C	T	A	A	C	G	T	G	A	G	C	T	C	C	A	A	C	C	A	A	C	A	C	C	A	C	A	C	C	C	C	A	G	C	A	G							
P4Ha	Cf	G	A	C	T	G	C	A	G	C	C	T	C	G	C	T	A	A	C	G	T	G	A	G	C	T	C	C	A	A	C	C	A	A	C	A	C	C	A	C	A	C	C	C	A	C	C	A	G	C	A	G					
Glut1	Hs	G	C	C	C	T	G	G	C	A	C	A	C	T	C	A	G	G	C	G	T	G	C	C	G	T	C	A	C	A	C	C	A	A	C	C	A	C	C	A	C	C	C	C	C	A	C	C	A	T	C	T					
Glut1	Mm	G	G	C	C	T	G	G	C	A	C	A	C	T	C	A	G	G	C	G	T	G	C	C	G	T	C	A	C	C	A	A	C	C	A	C	C	A	C	C	C	C	C	C	C	C	A	C	C	A	T	C	T				
RORa4	Hs	C	C	G	A	G	C	C	C	G	C	A	G	T	C	A	G	T	A	C	G	T	G	C	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	G	C	G				
RORa4	Cf	C	C	C	A	G	C	C	C	C	G	C	A	G	T	C	A	G	T	A	C	G	T	G	C	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	G	C	G			
RORa4	Cf	C	C	C	A	G	C	C	C	C	G	C	A	G	T	C	A	G	T	A	C	G	T	G	C	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	G	C	G		
EPO	Hs	G	A	G	G	G	C	T	G	G	C	C	C	C	C	T	A	A	C	G	T	G	C	C	T	C	A	C	A	C	C	C	A	A	C	A	C	C	A	C	C	A	C	A	G	C	C	C	C	T	C	T					
EPO	Cf	G	T	C	G	C	G	G	T	C	G	C	C	C	C	T	A	A	C	G	T	G	C	C	T	C	A	C	A	C	C	C	A	A	C	A	C	C	A	C	C	A	C	C	A	G	C	C	C	C	C	A	G	C	A		
PGK1	Hs	G	C	T	T	G	C	T	A	C	G	T	G	C	C	A	G	A	C	C	G	T	G	C	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
PGK1	Cf	A	G	A	C	T	C	A	C	A	G	T	G	C	A	T	A	A	C	C	G	T	G	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
VEGF	Hs	A	G	A	C	T	C	A	C	A	G	T	G	C	A	T	A	A	C	C	G	T	G	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
VEGF	Cf	A	G	A	C	T	C	A	C	A	G	T	G	C	A	T	A	A	C	C	G	T	G	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
TransfR	Hs	A	G	C	A	C	C	T	C	G	C	A	G	C	A	T	A	A	C	C	G	T	G	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
TransfR	Mm	G	C	A	C	T	C	G	C	T	A	G	T	G	T	A	A	C	C	G	T	G	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
TransfR	Mm	A	G	C	T	G	C	A	C	T	T	G	T	G	T	A	A	C	C	G	T	G	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
Transf	Hs	A	A	C	T	G	A	A	C	T	T	G	T	G	T	A	A	C	C	G	T	G	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Transf	Mm	C	C	C	G	C	A	G	T	G	C	C	A	A	T	A	A	C	C	G	T	G	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
DEC1	Hs	C	C	C	G	C	A	G	C	A	G	C	C	A	A	T	A	A	C	C	G	T	G	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
DEC1	Mm	C	C	C	G	C	A	G	C	C	A	G	C	C	A	T	A	A	C	C	G	T	G	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
DEC2	Hs	G	A	C	C	A	G	C	A	C	C	A	G	C	C	A	T	A	A	C	C	G	T	G	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
DEC2	Mm	G	A	C	C	A	G	C	A	C	C	A	G	C	C	A	T	A	A	C	C	G	T	G	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
DEC2	Mm	A	C	T	G	C	A	G	C	A	C	C	A	G	C	T	A	A	C	C	G	T	G	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
Met	Hs	A	C	T	G	C	A	G	C	A	C	C	A	G	C	T	A	A	C	C	G	T	G	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
Met	Mm	A	C	T	G	C	A	G	C	A	C	C	A	G	C	T	A	A	C	C	G	T	G	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
CA9	Hs	G	G	T	G	T	A	C	G	G	C	A	G	C	T	G	T	A	A	C	G	T	G	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
CA9	Mm	A	C	T	G	T	G	A	C	G	C	T	G	T	G	T	A	A	C	C	G	T	G	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
LDHA	Hs	C	G	T	G	G	A	G	C	C	A	G	C	C	G	G	A	A	C	C	G	T	G	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
LDHA	Mm	A	G	T	G	G	A	G	C	C	A	G	C	C	G	G	A	A	C	C	G	T	G	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
MIF	Hs	C	G	G	C	G	C	C	G	C	C	T	G	A	G	C	T	A	A	C	C	G	T	G	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
MIF	Mm	G	C	C	A	G	G	A	C	C	T	G	A	G	C	T	A	A	C	C	G	T	G	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	

ET1	Hs	T	A	G	T	C	A	C	C	A	A	C	A	G	C	A	C	G	T	G	A	C	G	T	G	C	A	G	C	A	T	A	A	G	G	C						
ET1	Mm	T	A	G	T	C	A	C	C	A	C	A	G	C	A	A	C	G	T	G	A	C	G	C	A	A	C	C	A	G	C	A	T	A	A	G	A	C				
Nur77	Hs	G	C	G	T	C	A	A	T	G	A	A	C	C	C	C	G	C	G	T	G	C	G	T	C	A	C	G	C	G	C	G	C	A	G	A	C					
Nur77	Mm	G	C	G	T	C	A	A	T	G	A	A	C	C	C	C	G	C	G	T	G	C	G	T	C	A	C	G	C	G	C	A	G	A	C							
Nur77_2HRE	Hs	C	C	T	G	A	A	T	G	T	C	T	G	C	G	C	A	C	G	C	A	C	C	A	C	G	C	A	C	G	C	G	G	T	T	C	C					
Nur77_2HRE	Mm	C	C	T	G	A	A	T	G	T	C	T	G	C	G	C	A	C	G	C	A	C	C	A	C	G	C	A	C	G	C	G	G	T	T	C	C					
CAD	Hs	G	C	G	G	A	C	C	C	C	G	C	C	C	T	A	C	G	T	G	C	C	G	C	C	C	G	C	C	C	C	C	C	C	C	T	C					
CAD	Mm	T	A	T	G	G	T	C	C	C	G	C	C	C	T	G	A	C	G	T	G	A	C	G	T	G	C	C	C	C	C	C	C	C	C	T	G					
A		9	9	4	4	5	9	9	14	4	4	5	10	4	4	7	2	38	0	0	0	0	0	0	0	0	10	8	4	7	21	9	9	11	9	23	4	15	11	9	7	5
C		13	23	18	15	13	14	19	16	25	16	22	6	18	25	13	10	0	46	0	0	0	0	0	0	32	17	10	15	11	18	9	16	23	11	20	8	13	16	21	15	
G		20	11	12	17	24	20	13	12	13	17	13	22	21	10	21	10	8	0	46	0	46	0	46	0	2	17	21	13	14	15	27	8	13	8	16	22	12	15	11	15	
T		4	3	12	10	4	3	5	4	4	9	6	8	3	7	5	24	0	0	0	46	0	0	0	2	4	11	11	0	4	1	11	1	4	6	1	10	6	7	11		
p value		0,4	0,1	0,1	0,3	0,1	0,3	0,7	0,1	0,0	0,5	0,3	0,0	0,1	0,0	0,4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,7	0,0	0,3	0,0	0,7	0,0	0,0	0,0	0,0	0,0	0,5	0,0	0,2	0,9	0,4	0,3	
Diferencia con la frecuencia esperada		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
P<0.05		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
P<0.01		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
P<0.001		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Entropía de Shanon (bits)		1,8	1,7	1,8	1,8	1,7	1,8	1,9	1,9	1,6	1,8	1,8	1,8	1,6	1,7	1,8	1,6	0,7	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,2	1,8	1,8	1,9	1,5	1,8	1,5	2,0	1,6	1,7	1,7	1,6	2,0	1,9	1,8	1,9	
Contenido de Información(bits)		0,1	0,2	0,0	0,0	0,2	0,1	0,0	0,0	0,3	0,1	0,1	0,1	0,3	0,2	0,1	0,2	1,2	1,9	1,9	1,9	1,9	1,9	1,9	0,7	0,1	0,1	-0,1	0,4	0,1	0,4	-0,1	0,3	0,2	0,1	0,3	-0,1	0,0	0,1	0,0		
Puntuación para cada base/posición		-0,9	1,7	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0	-6,0		
equilibrada por su contenido de información.		0,0	-3,7	3,6	-5,6	-5,6	-5,6	-5,6	-5,6	-5,6	-5,6	-5,6	-5,6	-5,6	-5,6	-5,6	-5,6	-5,6	-5,6	-5,6	-5,6	-5,6	-5,6	-5,6	-5,6	-5,6	-5,6	-5,6	-5,6	-5,6	-5,6	-5,6	-5,6	-5,6	-5,6	-5,6	-5,6	-5,6	-5,6	-5,6		
G		0,0	-0,4	-5,6	3,5	-5,6	3,5	-5,6	3,5	-5,6	3,5	-5,6	3,5	-5,6	3,5	-5,6	3,5	-5,6	3,5	-5,6	3,5	-5,6	3,5	-5,6	3,5	-5,6	3,5	-5,6	3,5	-5,6	3,5	-5,6	3,5	-5,6	3,5	-5,6	3,5	-5,6	3,5	-5,6	3,5	
T		0,3	-4,0	-6,0	-6,0	3,1	-6,0	-2,0	-0,3	0,0	-1,3	-0,3	-1,8	0,0	-1,5	-0,4	-0,3	-1,5	-0,4	-0,3	-1,5	-0,4	-0,3	-1,5	-0,4	-0,3	-1,5	-0,4	-0,3	-1,5	-0,4	-0,3	-1,5	-0,4	-0,3	-1,5	-0,4	-0,3	-1,5	-0,4		

## IX.ANEXO III

**Anexo III: Curva ROC.** Los genes utilizados como control positivo y negativo están marcados como "pos" y "neg". "TP" son los verdaderos positivos, "FP" los falsos positivos, "TN" los verdaderos negativos y "FN" los falsos negativos. "tp" = TP/(TP+FN) y "fp" = FP/(TP+FN). Queda señalado en rojo los valores de estos ratios para los dos puntos de corte a partir de los cuales aparecen diferencias significativas. Solo se muestran los primeros puestos del listado

marca	Gen	Pt/Pb	PUNTO DE CORTE DEL RATIO Pt/Pb											
			5,67						6,5					
			TP1	FP1	TN1	FN1	tp	fp	TP2	FP2	TN2	FN2	tp	fp
			0	0	0	0	0	0	0	0	0	0	0	0
pos	LDHA	3,18E+015	1	0	0	0	2,43902	0	1	0	0	0	2,43902	0
pos	VEGFA	3,67479E+13	2	0	0	0	4,87805	0	2	0	0	0	4,87805	0
pos	PGK1	3266033686	3	0	0	0	7,31707	0	3	0	0	0	7,31707	0
pos	EGLN1	16029121,73	4	0	0	0	9,7561	0	4	0	0	0	9,7561	0
pos	RORA	2440381,03	5	0	0	0	12,1951	0	5	0	0	0	12,1951	0
pos	P4HA1	836582,82	6	0	0	0	14,6341	0	6	0	0	0	14,6341	0
pos	MXI1	594440,93	7	0	0	0	17,0732	0	7	0	0	0	17,0732	0
pos	PLOD2	427892,47	8	0	0	0	19,5122	0	8	0	0	0	19,5122	0
pos	BNIP3	392078,73	9	0	0	0	21,9512	0	9	0	0	0	21,9512	0
pos	SLC2A1	60220,08	10	0	0	0	24,3902	0	10	0	0	0	24,3902	0
pos	MIF	11084,42	11	0	0	0	26,8293	0	11	0	0	0	26,8293	0
pos	DARS	8523,07	12	0	0	0	29,2683	0	12	0	0	0	29,2683	0
pos	PFKFB3	4158,86	13	0	0	0	31,7073	0	13	0	0	0	31,7073	0
pos	ENO1	3770,28	14	0	0	0	34,1463	0	14	0	0	0	34,1463	0
pos	PKM2	2032,65	15	0	0	0	36,5854	0	15	0	0	0	36,5854	0
pos	EGLN3	1678,07	16	0	0	0	39,0244	0	16	0	0	0	39,0244	0
pos	SERPINE1	1546,44	17	0	0	0	41,4634	0	17	0	0	0	41,4634	0
pos	GAPDH	1361,78	18	0	0	0	43,9024	0	18	0	0	0	43,9024	0
pos	PNRC1	1288,59	19	0	0	0	46,3415	0	19	0	0	0	46,3415	0
pos	CAD	1042,29	20	0	0	0	48,7805	0	20	0	0	0	48,7805	0
pos	PPP1R3C	506,71	21	0	0	0	51,2195	0	21	0	0	0	51,2195	0
pos	CITED2	329,36	22	0	0	0	53,6585	0	22	0	0	0	53,6585	0
pos	HK2	166,92	23	0	0	0	56,0976	0	23	0	0	0	56,0976	0
pos	PGAM1	60,1	24	0	0	0	58,5366	0	24	0	0	0	58,5366	0
pos	BCKDHA	49,45	25	0	0	0	60,9756	0	25	0	0	0	60,9756	0
pos	EDN1	46,65	26	0	0	0	63,4146	0	26	0	0	0	63,4146	0
pos	CA9	31,41	27	0	0	0	65,8537	0	27	0	0	0	65,8537	0
pos	GBE1	15,82	28	0	0	0	68,2927	0	28	0	0	0	68,2927	0
pos	TMEM45A	15,63	29	0	0	0	70,7317	0	29	0	0	0	70,7317	0
pos	RSBN1	11,44	30	0	0	0	73,1707	0	30	0	0	0	73,1707	0
neg	ZNRF2	10,93	30	1	0	0	73,1707	0,08258	30	1	0	0	73,1707	0,08258
neg	BMPR2	10,15	30	2	0	0	73,1707	0,16515	30	2	0	0	73,1707	0,16515
neg	HDAC2	9,91	30	3	0	0	73,1707	0,24773	30	3	0	0	73,1707	0,24773
neg	BCL2L11	9,91	30	4	0	0	73,1707	0,33031	30	4	0	0	73,1707	0,33031
neg	ARL2	9,63	30	5	0	0	73,1707	0,41288	30	5	0	0	73,1707	0,41288
neg	TEX264	9,19	30	6	0	0	73,1707	0,49546	30	6	0	0	73,1707	0,49546
neg	MAPK12	9,01	30	7	0	0	73,1707	0,57803	30	7	0	0	73,1707	0,57803
pos	SPRY1	8,55	31	7	0	0	75,6098	0,57803	31	7	0	0	75,6098	0,57803
neg	BAG3	8,25	31	8	0	0	75,6098	0,66061	31	8	0	0	75,6098	0,66061
neg	HOXA11	8,23	31	9	0	0	75,6098	0,74319	31	9	0	0	75,6098	0,74319
pos	UGP2	8,1	32	9	0	0	78,0488	0,74319	32	9	0	0	78,0488	0,74319
neg	TIAL1	7,99	32	10	0	0	78,0488	0,82576	32	10	0	0	78,0488	0,82576
neg	GNAI1	7,9	32	11	0	0	78,0488	0,90834	32	11	0	0	78,0488	0,90834
neg	SFR55	7,65	32	12	0	0	78,0488	0,99092	32	12	0	0	78,0488	0,99092
neg	CDH22	7,39	32	13	0	0	78,0488	1,07349	32	13	0	0	78,0488	1,07349
neg	TAF4B	7,35	32	14	0	0	78,0488	1,15607	32	14	0	0	78,0488	1,15607
neg	KIAA0427	7,28	32	15	0	0	78,0488	1,23865	32	15	0	0	78,0488	1,23865
neg	PLXND1	7,2	32	16	0	0	78,0488	1,32122	32	16	0	0	78,0488	1,32122
neg	FHOD3	7,13	32	17	0	0	78,0488	1,4038	32	17	0	0	78,0488	1,4038
neg	PAPPA	7,13	32	18	0	0	78,0488	1,48637	32	18	0	0	78,0488	1,48637
neg	EIF2C2	7,08	32	19	0	0	78,0488	1,56895	32	19	0	0	78,0488	1,56895
neg	ST6GAL1	7,08	32	20	0	0	78,0488	1,65153	32	20	0	0	78,0488	1,65153
neg	EXOSC5	7,03	32	21	0	0	78,0488	1,7341	32	21	0	0	78,0488	1,7341
neg	YAF2	6,84	32	22	0	0	78,0488	1,81668	32	22	0	0	78,0488	1,81668
neg	YAP1	6,79	32	23	0	0	78,0488	1,89926	32	23	0	0	78,0488	1,89926
neg	ACVR1B	6,74	32	24	0	0	78,0488	1,98183	32	24	0	0	78,0488	1,98183
neg	UVRAG	6,71	32	25	0	0	78,0488	2,06441	32	25	0	0	78,0488	2,06441
neg	COL8A1	6,69	32	26	0	0	78,0488	2,14699	32	26	0	0	78,0488	2,14699
neg	TF	6,57	32	27	0	0	78,0488	2,22956	32	27	0	0	78,0488	2,22956
neg	NRXN2	6,36	32	28	0	0	78,0488	2,31214	32	27	1	0	78,0488	2,22956
neg	RFX1	6,19	32	29	0	0	78,0488	2,39472	32	27	2	0	78,0488	2,22956
neg	LPHN1	6,19	32	30	0	0	78,0488	2,47729	32	27	3	0	78,0488	2,22956
neg	ACTN2	6,16	32	31	0	0	78,0488	2,55987	32	27	4	0	78,0488	2,22956
neg	ZFX	6,1	32	32	0	0	78,0488	2,64244	32	27	5	0	78,0488	2,22956
neg	ELF3	6,05	32	33	0	0	78,0488	2,72502	32	27	6	0	78,0488	2,22956
neg	VAPA	6,03	32	34	0	0	78,0488	2,8076	32	27	7	0	78,0488	2,22956
neg	CAB39	5,96	32	35	0	0	78,0488	2,89017	32	27	8	0	78,0488	2,22956
neg	AMOTL2	5,88	32	36	0	0	78,0488	2,97275	32	27	9	0	78,0488	2,22956
neg	ZFY	5,76	32	37	0	0	78,0488	3,05533	32	27	10	0	78,0488	2,22956



X.ANEXO IV

**Anexo IV. Genes diana de HIF predichos por nuestra estrategia bioinformática.** Posición HBS, coordenadas de los sitios RCGTG. N\_ especies, Número de especies en los que está conservado ese HBS. Cuando existe más de un HBS por gen se muestra separado por barras verticales. Sondadas, número de sondas para ese gen que aparecen en todos los perfiles de GEO. Media y SD, son los valores asociados a la fluorescencia de las sondas usadas para el cálculo de los valores LOG2(hypoxia/normoxia). pvalor\_ajustado, es el pvalor ajustado por FDR para el test de comparación con respecto a 0. pT, probabilidad de ser un gen diana de HIF, P(x,y,p | T); Pn, probabilidad de no ser una diana de HIF, P(x,y,p | B). pT/pB es el ratio de P(x,y,p | T)/P(x,y,p | B). Diana conocida, los genes se consideran como conocidos si aparecen como conocidos en la revisión de (Wenger, Stiehl et al. 2005) (STKE) y si existe al menos una publicación que lo relacione con HIF e hipoxia(PMID indicado en la columna REF)

Chr	Gen	Posición	N_ especies	Secuencia HBS	Puntuación	Máx HBS	Sondadas	Media	SD	pvalor_ajustado	pT	pB	pT/pB	Diana conocida	REF	STKE
chr11	LDHA	183725948 17		CACGTGTGAGTC	12.95 18.32 18.32	69	2,28	1,50	1,00	3,30E-015	1,00	0,00	2,89E+019	SI	PMID: 8955077	SI
chr6	VEGFA	43844944 13 11 9 15		TACGTGGGCTCC	15.12 11.98 18.23	47	3,43	1,76	1,00	2,79E-013	1,00	0,00	2,79E+017	SI	PMID: 8756616	SI
chr8	BNIP3L	26296354 21 7		GACGTGGAGACG	13.90 17.62 17.62	35	3,66	1,77	1,00	4,13E-010	1,00	0,00	8,59E+013	SI	PMID: 10922063	SI
chrX	PGK1	77246198 9 17		GACGTGACAAAC	15.89 13.25 15.89	60	2,14	1,73	1,00	5,38E-010	1,00	0,00	1,25E+013	SI	PMID: 8089148	SI
chr14	ERO1L	5,2E+07	7	GACGTGGCGG	18,17	34	3,44	2,2	1,00	2,92E-007	1,00	0,00	4,77E+010	SI	PMID: 12752442	SI
chr1	EGLN1	2,3E+08	5	TACGTGCAGAGC	18,06	33	3,02	1,96	1,00	6,01E-007	1,00	0,00	1,76E+010	SI	PMID: 14506252	SI
chr15	RORA	58576248 27 15 9 7 9		AGCGTGTTCAC	10.81 19.01 19.02	34	2,28	1,71	1,00	4,77E-006	1,00	0,00	1,42E+009	SI	PMID: 15270719	SI
chr10	P4HA1	74525338 9 16 9		GACGTGGCAGCA	13.24 16.64 16.65	22	3,16	1,54	1,00	4,66E-006	1,00	0,00	5,71E+008	SI	PMID: 14622280	SI
chr10	MXI1	11195798 43 16 12 2		TGCGTGTCTTTC	11.01 10.85 15.96	21	3,63	1,63	1,00	3,21E-006	1,00	0,00	4,21E+008	SI	PMID: 16319523	SI
chr3	PLOD2	1,5E+08	12	TACGTGTGGAGG	15,4	28	2,58	1,54	1,00	2,19E-006	1,00	0,00	3,26E+008	SI	PMID: 14622280	SI
chr10	PFKP	3101241	6	TACGTGATGCC	17,29	27	1,47	1,47	1,00	1,15E-005	1,00	0,00	3,09E+008	SI	PMID: 17193925	SI
chr19	GADD45E	2426956 9 20		CACGTGGGGCC	13.81 12.08 13.82	44	1,68	1,24	1,00	4,66E-008	1,00	0,00	2,95E+008	NO		
chr1	AK3L1	65386148 6 15		CACGTGGCAGGA	16.38 10.98 18.19	26	3,82	2,45	1,00	1,59E-005	1,00	0,00	2,74E+008	NO		
chr10	BNIP3	133645545 10 15		CACGTGGCGGC	18.00 14.17 18.01	26	4,15	2,75	1,00	2,41E-005	1,00	0,00	1,46E+008	SI	PMID: 10922063	SI
chr9	NFL3	93224347 7 15 29		GCGTGGAGTTA	9.44 10.02 17.85	21	2,22	1,23	1,00	3,83E-005	1,00	0,00	8,17E+007	NO		
chr5	LOX	12144088 10 8 21 30		GACGTGGCTGCC	12.60 10.92 15.98	34	3,38	2,82	1,00	2,35E-005	1,00	0,00	3,12E+007	SI	PMID: 16642001	SI
chr1	SLC2A1	4,3E+07	7	GACGTGGCGTC	15,73	28	3,22	2,28	1,00	2,40E-005	1,00	0,00	2,20E+007	SI	PMID: 15525582	SI
chr5	STC2	17267735 7 9 7 23		TGCGTGTGACAT	11.78 14.30 16.49	26	3,3	2,44	1,00	9,93E-005	1,00	0,00	9,22E+006	SI	PMID: 18394600	SI
chr8	STC1	2,4E+07	10	TACGTGCCAGAT	16,1	34	2,63	2,43	1,00	1,07E-004	1,00	0,00	5,46E+006	SI	PMID: 19628018	SI
chr8	KLF10	1037354 14 24 30 37		GACGTGGCAGCC	18.09 12.61 18.09	22	1,32	0,86	1,00	1,43E-004	1,00	0,00	4,13E+006	NO		
chr22	MIF	2,3E+07	7	TACGTGCCCTGAC	16,26	18	1,48	0,83	1,00	2,40E-004	1,00	0,00	2,11E+006	SI	PMID: 16854377	SI
chr10	DDIT4	73703562 20 28		TACGTGGCCCCG	17.62 11.64 17.63	19	3,26	2,27	1,00	1,12E-003	1,00	0,00	9,46E+005	SI	PMID: 17307335	SI
chr3	SLC25A3	1,4E+08	7	CACGTATCCCTA	14,32	53	1,41	0,94	1,00	3,40E-011	1,00	0,00	6,69E+005	NO		
chr10	PKFB3	6245764 6 12 8 16 7		GCGTGGCCGAC	16.94 16.90 16.94	23	2,95	2,46	1,00	1,20E-003	1,00	0,00	5,33E+005	SI	PMID: 11744734	SI
chr10	ENO1	8861690 6 12 18		TACGTGACGGAG	16.74 13.67 16.74	30	1,45	1,48	1,00	1,12E-003	1,00	0,00	5,08E+005	SI	PMID: 8955077	SI
chr10	HK1	70748307 5 11 17		CGCGTGGCGGC	11.43 15.38 16.38	20	2,47	1,81	1,00	1,15E-003	1,00	0,00	3,46E+005	SI	PMID: 16438736	SI
chr11	ADM	10282383 15 13 10 20		AACGTGAGGACT	14.90 12.32 18.17	19	5,15	4,01	1,00	2,89E-003	1,00	0,00	3,23E+005	SI	PMID: 1515277	SI
chr12	BTG1	91062854 11 15 21		GACGTGGCTCT	17.04 12.86 17.05	37	1,22	1,35	1,00	4,87E-004	1,00	0,00	2,93E+005	SI	PMID: 19068236	SI
chr2	CXCR4	13659175 9		AGCGTGCAAGCC	14.93 6.71	33	2,59	2,52	1,00	2,77E-004	1,00	0,00	2,87E+005	SI	PMID: 15235597	SI
chr11	PMVE1	7,4E+07	8	TACGTGACCATA	14,57	24	1,38	0,95	1,00	9,66E-005	1,00	0,00	1,81E+005	NO		
chr18	SMAD7	44279017 9		CACGTGGAGGA	14.14 15.53 15.53	21	1,29	0,91	1,00	5,17E-004	1,00	0,00	2,62E+005	NO		
chr7	SERPINE1	1E+08	7	CACGTGAGGGG	16,56	28	1,74	1,82	1,00	2,32E-003	1,00	0,00	1,60E+005	SI	PMID: 11877282	SI
chr1	PLOD1	119126319 18 9		CACGTGTGTGG	14.41 18.16 18.16	18	1,31	1,03	1,00	4,73E-003	1,00	0,00	1,56E+005	SI	PMID: 14622280	SI
chr6	PNRC1	9E+07	9	TACGTGCCCGC	17,25	19	1,43	1,19	1,00	4,87E-003	1,00	0,00	1,01E+005	NO		
chr14	EGLN3	3,3E+07	13	TACGTGGCTGCK	18,87	24	3,15	3,33	1,00	6,89E-003	1,00	0,00	9,96E+004	SI	PMID: 15823097	SI
chr9	NR4A3	1016213 8 20 26 81		TACGTGGAACCT	14.45 13.36 17.96	27	1,16	1,28	1,00	4,98E-003	1,00	0,00	7,52E+004	NO		

chr6	ZNF292	879219249 9 20 31 42	TACGTGCTCGGC	18.08 16.30 18.09	26	1.42	1.64	8,46E-003	1.00	0,00	6,56E+004	NO	
chr10	AVP1	994370616 18 19	GGCGTGTCCATC	10.48 11.41 15.35	16	1.29	0,76	1,20E-003	1.00	0,00	6,26E+004	NO	
chr3	YEATS2	1,8E+08 7	CACGTGCCCGC	16,99	22	1,63	1,62	7,06E-003	1,00	0,00	4,82E+004	NO	
chr6	RRAGD	901778409 7 18	GGCGTACTGGC	12.61 14.85 14.89	26	1,75	1,61	1,17E-003	1,00	0,00	4,24E+004	NO	
chr10	PPP1R3C	9,3E+07 9	CACGTGCTGGG	17,12	18	2,01	1,82	1,13E-002	1,00	0,00	2,66E+004	NO	
chrX	PKD3	243932628 5	TACGTGCAGCA	16.21 14.48 16.21	37	1.13	0.98	1,44E-005	1,00	0,00	2,48E+004	SI	PMID: 18718909
chr2	IGFBP5	2,2E+08 5	TGCGTGACCTG	14,45	51	1,4	2,03	9,84E-004	1,00	0,00	2,36E+004	SI	PMID: 7684482
chr2	DARS	1,4E+08 7	GACGTGCTAAC	16,63	32	1,09	0,83	1,46E-005	1,00	0,00	1,75E+004	NO	
chr6	CITED2	13973689 21 34 10	CACGTGTGGCG	13.08 12.73 18.53	30	1,3	2,01	3,35E-002	1,00	0,00	1,03E+004	SI	PMID: 16740701
chr2	ABCB6	21979178 7 15 22 30	CACGTGCTCC	16.15 14.54 16.15	21	1,09	0,97	4,08E-003	1,00	0,00	8733,60	NO	
chr2	HK2	74914395 6 11 18 28	GGCGTGTCTGG	10.58 13.61 15.98	23	2,78	3,07	1,17E-002	1,00	0,00	8342,64	SI	PMID: 17193925
chr5	P4HA2	13159150 7 18	GACGTGGCGGC	12.90 10.05 12.91	21	2,75	1,75	1,80E-004	1,00	0,00	6786,29	SI	PMID: 14622280
chr22	MAFF	36927750 5 16	CACGTGACACT	14.89 13.58 14.9	22	2,77	2,62	4,95E-003	1,00	0,00	5948,53	SI	PMID: 16580664
chr15	PKM2	70306270 8 17	TACGTGCAGACA	17.63 15.30 17.64	21	0,99	0,76	1,23E-003	1,00	0,00	5836,11	SI	PMID: 9692838
chr1	EFNA3	15331793 7 17 32 39	GGCGTGGCTAA	12.88 11.52 16.95	17	1,25	1,3	3,35E-002	1,00	0,00	5006,31	NO	
chr1	RLF	4E+07 8	CACGTGACTCAG	16,06	18	1,53	1,56	2,42E-002	1,00	0,00	3334,72	NO	
chr12	GAPDH	6512952 7 15 9 18	GACGTGCAGGC	16.98 15.46 16.98	89	0,97	0,6	1,48E-021	1,00	0,00	2732,16	SI	PMID: 7929107
chr19	JUNB	12763149 9 18 12 22	CGCGTGTCTTG	8.60 12.62 12.87	21	1,34	0,87	2,22E-004	1,00	0,00	2286,41	SI	PMID: 16869749
chr10	RP11-529	1E+08 9	CACGTGACTCT	13,35	18	1,51	0,99	9,89E-004	1,00	0,00	2252,56	NO	
chr6	FOXO3	1089858 7 15 22 27	CGCGTGGCGGC	12.76 8.20 17.93	48	0,92	1,1	1,13E-004	1,00	0,00	2066,55	SI	
chr15	CHSY1	1E+08 17	CACGTGAACCC	15,54	18	1,04	0,74	2,17E-003	1,00	0,00	1825,61	NO	
chr3	GBE1	8,2E+07 13	AGCGTACTGGA	12,48	18	2,8	1,55	2,16E-004	1,00	0,00	1659,71	SI	
chr6	HIVEP2	14314743 7 6 11 22 3	GACGTGGTTCT	10.78 11.88 17.25	30	0,91	1,06	3,96E-003	1,00	0,00	1228,07	NO	
chr7	HOXA4	27137095 8 13	CACGTGGGCAC	15.58 11.03 15.58	19	1,03	1,01	1,42E-002	1,00	0,00	1173,06	NO	
chr10	PGAM1	9,9E+07 8	TACGTGCGCTCC	16,98	8	1,4	0,9	8,57E-002	1,00	0,00	1166,89	SI	PMID: 19068236
chr8	TPD52	8,1E+07 6	GACGTGGCGGA	15,08	61	1,05	0,81	8,68E-011	1,00	0,00	1129,24	NO	
chr3	TRPC1	1,4E+08 7	GACGTGAGACA	16,32	29	0,95	0,83	2,40E-004	1,00	0,00	1048,75	SI	PMID: 16709899
chr22	PDGFB	37961251 9 19 7 18 3	CACGTGGCGGC	14.65 17.75 17.8	30	0,89	1,43	4,06E-002	1,00	0,00	1020,21	SI	PMID: 16136272
chr5	AP3S1	11520552 5 19	CGCGTGGCGGA	16.41 18.93 16.42	20	0,94	0,78	3,12E-003	1,00	0,00	977,68	NO	
chr12	APOLD1	12767905 8 15	GACGTGCTGGC	18.31 11.76 18.32	11	1,2	1,28	1,45E-001	1,00	0,00	952,82	NO	
chr6	NRN1	5943717 6 7 14 16 12	GACGTGAGCTG	9.99 9.84 11.81	13	2,61	2,39	5,23E-002	1,00	0,00	756,25	SI	PMID: 16723126
chr1	KIAA2013	11912631 9 18	CACGTGCTGG	14.41 18.16 18.16	23	0,85	0,7	1,16E-003	1,00	0,00	751,17	NO	
chr6	DDR1	30957991 6 15	TACGTGCTCTC	16.80 10.10 16.81	56	0,89	0,9	6,41E-007	1,00	0,00	642,77	NO	
chr8	PSD3	18432375 7 120	AACGTGTGCTTT	10.74 13.70 16.33	36	0,92	1,33	8,46E-003	1,00	0,00	632,63	NO	
chr3	TMEM45	10169420 7 14	GACGTGCTCGC	13.19 14.61 14.62	15	4,48	3,75	1,83E-002	1,00	0,00	551,66	SI	PMID: 17185612
chr9	CA9	3,6E+07 9	TACGTGCAATGG	16,77	13	2,72	3,29	1,38E-001	1,00	0,00	452,48	SI	PMID: 14712082
chr19	APIP1	41051203 5 19	AGCGTGGCGGC	8.86 15.67 15.68	15	0,99	0,99	4,40E-002	1,00	0,00	437,32	NO	



chr3	SIH2	15196235 8 7 14	GGCGTGGCTTGC 9.72 15.08 15.09	18	1	0.92	1,24E-002	1,00	0,00	384,69	NO	
chrX	TSC22D3	1,1E+08 7	TGCGTGCCGTTTCATCC	33	1,17	1,47	4,34E-003	1,00	0,00	372,6	NO	
chr10	SFXN3	1E+08 8	CGCGTGAGCGCG 13,84	31	1,23	1,59	7,53E-003	1,00	0,00	369,74	NO	
chr8	PAG1	8,2E+07 6	GACGTGCCGCTG 16,1	23	0,91	1,24	4,28E-002	1,00	0,00	329,92	NO	
chr2	SCN1A	16655498 6	AACGTGCACCAC 15.50 8.64 15.5	11	0,95	0,6	2,20E-002	1,00	0,00	312,53	NO	
chr11	BRSK2	1367769 6 4 10	GGCGTGCGCTGG 15.53 16.58 16.59	9	1,08	1,02	1,66E-001	1,00	0,00	268,62	NO	PMID: 17554077
chr6	EDN1	1,2E+07 9	AACGTGCAGCGG 16,7	24	0,85	1,28	6,19E-002	1,00	0,00	262,75	SI	
chrX	PCYT1B	24489364 7 5 7	TACGTGCAGGTA 16.26 10.32 16.27	23	0,88	1,31	6,62E-002	1,00	0,00	238,48	NO	
chr2	KLF7	2076829 8 9 8 14 6	CACGTGACATCT 14.08 14.37 15.2	34	0,95	1,14	2,32E-003	1,00	0,00	235,78	NO	
chr6	PIM1	37244638 6 15 9	GACGTGTGAAAT 13.07 14.87 14.88	21	1	1,15	2,37E-002	1,00	0,00	213,53	SI	PMID: 18708761
chr11	PHF21A	46004880 11 7 21 29	AACGTGCACAGA 14.17 17.03 17.04	35	0,8	0,91	1,11E-003	1,00	0,00	204,15	NO	
chr4	SPRY1	12453786 12 20	TGCGTGTGTTT 8.12 15.10 15.1	17	1,07	1,31	7,07E-002	0,99	0,01	152,21	NO	
chrX	TCEAL3	10274948 7 14	CACGTGAGCGCC 16.10 16.05 16.1	10	0,87	0,74	9,54E-002	0,99	0,01	147,77	NO	
chr11	VEGFB	6,4E+07 5	TACGTGCGGCC 18,78	20	0,72	0,82	2,69E-002	0,99	0,01	134,52	SI	PMID: 11303624
chr8	PLEKHA2	3,9E+07 7	AACGTGAGAAAA 12,91	23	1,86	1,75	3,64E-003	0,99	0,01	130,47	NO	
chr3	PCOLCE2	1,4E+08 6	CACGTGGTGGCC 13,8	16	1,08	0,82	7,53E-003	0,99	0,01	118,76	NO	
chr6	TNFAIP3	13822957 6 10 6	CACGTGCACAGC 16.05 13.31 16.06	27	0,85	1,53	9,60E-002	0,99	0,01	113,73	SI	PMID: 19536297
chr15	FAM63B	5,7E+07 7	GACGTGCCTGAC 16,64	22	0,78	0,69	2,95E-003	0,99	0,01	110,41	NO	
chr1	PLCH2	2393493 5	GACGTGCGGGG 18,47	6	1,35	1,69	4,67E-001	0,99	0,01	109,99	NO	
chr12	CDKN1B	12761388 8 8	CGCGTGACTGT 13.69 10.62 13.7	22	1,07	0,88	1,53E-003	0,99	0,01	109,48	SI	PMID: 19342889
chr19	BCKDHA	4,7E+07 13	CACGTGCGGACG 18,15	21	0,71	0,84	2,79E-002	0,99	0,01	109,16	NO	
chr14	MIDGA2	46380435 9 12 8 22	TACGTGAAGAGC 12.13 10.92 17.9	4	2,29	1,54	4,50E-001	0,99	0,01	103,39	NO	
chr7	EGFR	5,5E+07 7	CACGTGCGCCCC 17,88	74	0,7	2,45	1,30E-001	0,99	0,01	81,93	SI	PMID: 16540671
chr2	YPEL5	30222936 10 4 8	CACGTGAGTCAG 12.25 12.07 16.44	26	0,77	0,71	1,16E-003	0,99	0,01	78,85	NO	
chr10	KLF6	3815288 10 21 7 14	TGCGTGAAGAA 13.63 11.19 14.83	58	0,92	1,12	2,19E-005	0,99	0,01	78,48	NO	
chr5	HOMER1	78844824 7 29 43 51	GACGTGCACGTG 13.43 13.05 16.04	27	0,8	1,05	1,71E-002	0,99	0,01	75,95	NO	
chr3	RYBP	7,3E+07 20	CGCGTGCCCTGCGTGGC 12,5	38	1,2	0,9	9,39E-007	0,99	0,01	74,5	NO	
chr20	PHACTR3	57613110 14 6	TGCGTGCTTGGG 15.60 13.48 15.6	6	0,84	0,22	2,99E-002	0,99	0,01	73,12	NO	
chr9	KLF4	10929207 14	CACGTGCGCCGA 14.25 12.71 14.25	23	1,08	1,46	4,12E-002	0,99	0,01	70,26	NO	
chr19	GYS1	5,4E+07 14	GACGTGTGGTTA 12,52	20	2,15	1,76	2,90E-003	0,98	0,02	62,76	NO	
chrX	ZDHHC9	12880511 12 16 20 26	GGCGTGGCGGAA 10.47 8.89 15.19	13	0,86	0,61	1,54E-002	0,98	0,02	54,84	NO	
chr5	SPOCK1	13637495 6 11	AACGTGCCTTGA 14.82 15.00 15	9	2,23	1,9	1,29E-001	0,98	0,02	53,8	NO	
chr5	FNIP1	13116010 7 7 7	GGCGTGTGGACC 10.20 13.59 15.71	29	0,8	1	8,71E-003	0,98	0,02	50,04	NO	
chr5	FOXD1	7,3E+07 8	GACGTGCACACC 16,24	20	0,76	1,23	1,32E-001	0,98	0,02	46,83	NO	
chr12	PTPRR	6,9E+07 7	TACGTGCTGTTCT 15,15	21	0,85	0,8	6,33E-003	0,98	0,02	44,98	NO	
chr2	KCNGB3	4,3E+07 8	TACGTGGGTCCG 15,18	4	1,17	0,28	1,69E-001	0,98	0,02	44,64	NO	
chr17	MINT	2234606 7 9 17 11 9	TGCGTGTTTTACC 9.68 13.98 16.61	24	0,71	0,89	2,14E-002	0,98	0,02	42,67	NO	

chr3	BCL6	18893648 13 10 18	TACGTGCTGTAG	17.69 13.71 17.69	33	0.64	1.04	2.87E-002	0.98	0.02	40.29	NO	
chr18	RNF165	42167168 9 17 25 3	GGCGTGTCTGTG	10.79 12.64 17.86	14	0.65	1.36	3.87E-001	0.97	0.03	37.62	NO	
chrX	BCOR	398031347 10 8 6 12	CACGTGACAGGG	13.19 9.70 17.46	38	0.64	0.97	8.78E-003	0.97	0.03	35.86	NO	
chr7	CAV1	1.2E+08 6	GACGTGTCTGCTGC	15.32 15.32	26	0.82	1.22	4.38E-002	0.97	0.03	35.37	NO	
chr1	FBXO42	1.7E+07 11	GACGTGAGTGGC	16.18 16.18	34	0.73	0.77	6.27E-004	0.97	0.03	35.33	NO	
chr10	RHOB1	62372913 7	AACGTGCAAGC	15.38 17.95 17.96	16	0.64	1.54	4.12E-001	0.97	0.03	35.18	NO	
chr17	FGF11	7282705 13 9 18 26	AACGTGGGTGCA	12.60 16.61 16.62	9	0.72	0.88	2.76E-001	0.97	0.03	34.66	NO	
chr11	WEE1	9551285 7 11 18	GGCGTGGCAGT	11.31 15.02 17.42	27	0.64	1.08	7.00E-002	0.97	0.03	34.51	NO	
chr3	ADCY5	12448543 8 6 12 17 2	TGCGTGTAAAG	10.42 17.84 17.85	5	0.68	1	6.03E-001	0.97	0.03	32.54	NO	PMID: 15252039
chr2	ID2	8736333 6 12 19 26	GACGTGCAGCGT	16.54 11.68 16.55	19	0.69	1.02	1.09E-001	0.97	0.03	29.95	SI	
chr11	DSCAML1	1116888345 9 15 4 13	AGCGTGTATCATA	10.71 7.73 18.51	8	0.61	1.1	5.07E-001	0.97	0.03	29.7	NO	
chr10	XPNEP1	111673159 18	GACGTGCCCCAG	18.30 15.83 18.31	34	0.59	0.88	1.60E-002	0.97	0.03	28.96	NO	
chr5	DUSP1	172130587 7	AGCGTGTATCGGC	10.99 12.87 12.87	26	1.61	1.9	9.95E-003	0.97	0.03	27.58	SI	PMID: 16081065
chr3	RAP2B	154362745 11	AACGTGCCCCA	16.47 12.44 16.47	40	0.68	1.01	6.16E-003	0.96	0.04	26	NO	
chr14	ZBTB25	64040217 14 34 5 5	TGCGTGAAGCCG	13.42 12.71 13.42	18	1.12	1.15	2.45E-002	0.96	0.04	25.39	NO	
chr16	BCAR1	73830224 10 6	GACGTGGCAGGT	11.61 16.76 16.77	16	0.66	1.1	2.11E-001	0.96	0.04	24.44	SI	PMID: 17001309
chr3	OXSRI1	381815048 11	GACGTGCCAAAAG	15.74 11.13 15.74	19	0.74	0.52	1.18E-003	0.96	0.04	23.42	NO	
chr11	SERGF	1.8E+07 6	TACGTGACTCCG	16.72 16.72	24	0.65	1.09	9.67E-002	0.96	0.04	22.55	NO	
chr3	IL1RAP	191734148 8	CACGTGCTCTGG	14.48 10.10 14.48	28	0.88	1.07	8.47E-003	0.96	0.04	22.1	NO	
chr1	RSBN1	114156709 18	AGCGTGCAGAA	15.79 13.42 15.79	50	0.73	1.03	8.23E-004	0.96	0.04	21.99	NO	
chr6	GJA1	12179855 10	GGCGTAGGAA	14.17 11.16 14.18	25	1.62	2.78	9.42E-002	0.95	0.05	19.19	NO	
chr15	CLK3	72701311 13 8	GACGTGCCCTT	12.45 9.88 12.45	22	1.2	1.13	4.71E-003	0.95	0.05	19.1	NO	
chr2	BCL2L11	111594746 8 9	GGCGTGTACC	11.11 18.16 18.16	63	0.01	1.29	9.91E-001	0.95	0.05	18.86	NO	
chr2	DYSF	71617174 11 8 15 9 8	TGCGTGTCTCTG	7.03 14.56 17.5	11	0.58	0.74	2.17E-001	0.95	0.05	18.78	NO	
chr22	ATF4	38246429 6 9	CACGTGGTCGGC	13.71 14.45 14.45	21	0.87	0.58	2.77E-004	0.95	0.05	18.12	SI	PMID: 19519398
chr3	RASSF1	5E+07 6	TGCGTGCCTGT	13.98 13.98	20	1.03	1.41	6.88E-002	0.95	0.05	17.8	NO	
chr2	BMPR2	202948116 16	CGCGTGGGAGG	9.43 18.46 18.46	49	0.03	0.9	9.25E-001	0.95	0.05	17.18	NO	
chr3	GRAMD1	1.2E+08 5	CACGTGCGGACA	18.39 18.39	14	0.04	0.65	9.26E-001	0.94	0.06	17.03	NO	
chr12	DDN	47678844 10 12	CACGTGCACAG	17.41 10.54 17.42	9	0.58	1.03	4.55E-001	0.94	0.06	16.91	NO	
chr7	FLJ23834	1.1E+08 6	GACGTGAATAAAA	16.02 16.02	3	1.95	0.79	5.29E-001	0.94	0.06	16.1	NO	
chr7	IGFBP3	4.6E+07 8	GGCGTCTCTGG	12.96 12.96	25	3.43	4.25	1.71E-002	0.94	0.06	15.79	SI	PMID: 17486380
chr1	MEGF6	3504959 7	TACGTGCTGCTT	15.81 15.81	13	0.88	2.04	4.61E-001	0.94	0.06	15.37	NO	
chr19	MKNK2	1989680 7	GACGTGGCTACG	15.27 15.27	26	0.76	0.96	1.60E-002	0.94	0.06	15.21	NO	
chr3	TEX264	516802047 8	CGCGTGCAGCG	15.24 18.24 18.24	25	0.04	0.55	8.83E-001	0.94	0.06	14.97	NO	
chr1	NEGR1	71645624 19 9 8 7	AGCGTGGGATAT	8.50 8.93 11.56 12	15	0.75	1.31	2.59E-001	0.94	0.06	14.54	NO	
chr9	IERSL	130978639 7	CACGTGAGGGAG	16.44 14.43 16.45	14	0.69	1.94	5.24E-001	0.94	0.06	14.48	NO	
chr19	ZNF653	114544837 15 22	CGCGTGGCGCCA	11.56 17.39 17.39	7	0.56	0.83	4.77E-001	0.94	0.06	14.42	NO	

chr2	UGP2	639224438 8		GCGTGGGGACT 10.11 14.81 14.81	32	0.81	0.93	2,27E-003	0.94	0.06	14,4	NO	
chr20	CDH22	44263038 14 10 18		CGCGTAAAAAG 10.98 17.54 17.54	16	0.01	0.79	9,79E-001	0.93	0.07	14,31	NO	
chr7	HOXA11	27190414 23 15 12 12		CACGTGGCGAGT 12.48 18.70 17.92	21	0.07	0.94	8,90E-001	0.93	0.07	13,8	NO	
chr15	DL4	3900550 12 20 30 40		CGCGTGCCAACT 14.93 11.25 18,09	7	0.12	0.61	8,60E-001	0.93	0.07	13,71	SI	PMID: 17045587
chr3	PLXND1	13077908 15		CACGTGCCGGGC 17.54 13.05 17,54	29	0.03	1.01	9,50E-001	0.93	0.07	13,36	NO	
chr14	SFR5	6930386 11 8		GACGTGCCCCAG 17.74 12.37 17,75	39	0.04	0.77	8,97E-001	0.93	0.07	13,15	NO	
chr3	CCNL1	15836057 13 19 25		TGCGTGGGGCTC 13.42 14.06 14,06	34	0.9	1.24	7,68E-003	0.93	0.07	13,03	NO	
chr4	MTPP	1E+08 9		TAGTGCCCTCC 15.93	11	0.81	1.92	5,22E-001	0.93	0.07	12,67	NO	
chr19	BBC3	52416206 6 37 46		CACGTGTGACCA 13.33 13.47 15,23	17	0.75	0.83	4,46E-002	0.92	0.08	12,3	SI	PMID: 17998337
chr12	ACVR1B	5,1E+07 5		CACGTGCGCTCC 17,45	51	0.03	1.02	9,37E-001	0.92	0.08	12,28	NO	
chrX	ZFX	240774336 13 20 25		CGCGTGTGTGGC 12.55 14.20 17,24	49	0.01	0.64	9,75E-001	0.92	0.08	11,77	NO	
chr9	PAPPA	118053447 5 8 7 11		AGCGTGACGTTA 12.36 17.70 17,7	34	0.06	0.85	8,51E-001	0.92	0.08	11,53	NO	
chr6	ELOVL4	8,1E+07 7		CACGTGACGCCG 14,69	12	0.83	0.88	1,12E-001	0.92	0.08	11,03	NO	
chr7	GNAI1	8E+07 5		CACGTGCCCTCG 18,16	26	0.14	1.07	7,60E-001	0.92	0.08	11,03	NO	
chr9	FIBCD1	13277897 9		TGCGTCTCAGG 16.10 17.31 17,31	9	0.12	1.19	9,18E-001	0.92	0.08	10,86	NO	
chr11	AHNAK	62024568 5		TACGTGCTGCC 17.77 13.90 17,77	34	0.5	1.27	1,91E-001	0.91	0.09	10,64	SI	PMID: 17437048
chr18	DTNA	30327019 18 12 20		CACGTGACGAG 15.41 15.53 15,54	61	0.7	1.44	1,12E-002	0.91	0.09	10,59	NO	
chr1	ELF3	2E+08 6		CACGTGCTCCCTG 17,36	25	0.1	1.38	8,82E-001	0.91	0.09	10,33	NO	
chr5	DDX4	5,5E+07 11		CACGTGCAAGAC 17,13	9	0.05	0.87	9,49E-001	0.91	0.09	10,31	NO	
chrY	ZFY	2863212 10 17 24 31		GGCGTGACCGGC 13.80 14.20 17,24	17	0.09	1.28	9,13E-001	0.91	0.09	10,25	NO	
chr11	NRXN2	641305338 16 8 8 8		GACGTGATTGA 14.84 17.45 17,5	16	0.1	0.85	8,45E-001	0.91	0.09	10,22	NO	
chr7	LOC4025	1E+08 8		CACGTGCTGGC 17,99	4	0.31	0.46	6,99E-001	0.91	0.09	10,16	NO	
chr15	IGF1R	97007828 8 11 16 21		TACGTGGACTTT 10.67 13.95 18,24	52	0.48	0.92	1,31E-002	0.91	0.09	10,12	SI	PMID: 19016246
chr16	GPT2	4,5E+07 7		AACGTGCTGTGCCCT	13	0.83	0.9	9,67E-002	0.91	0.09	10,04	NO	
chr7	NXPH1	8440812 5 15 7 14		GACGTGCCCTCC 17.41 14.02 17,41	5	0.2	0.71	8,47E-001	0.91	0.09	9,95	NO	
chrX	PIM2	486611037 15 7		GGGTGGACGCC 11.29 13.98 13,99	16	0.89	0.54	1,37E-003	0.91	0.09	9,67	NO	
chr20	PHF20	33823207 6 14 28 3		GGCGTGGCCTGC 12.13 11.09 17,05	54	0.01	0.57	9,48E-001	0.91	0.09	9,65	NO	
chr11	RCOR2	634406338 13 23 29		AGCGTGGTAGG 12.43 10.46 18,32	8	0.44	0.62	3,99E-001	0.91	0.09	9,63	NO	
chr20	BMP2	6697478 16 29 10 7		GACGTGTCCCG 10.68 15.36 15,37	22	0.74	1.61	2,40E-001	0.91	0.09	9,62	SI	PMID: 12045566
chr12	TBK1	6,3E+07 11		CACGTGACCGAC 17,31	23	0.06	0.65	8,64E-001	0.91	0.09	9,54	SI	PMID: 16537515
chr5	JMY	78567657 14		AGCGTGGGCAC 15.12 15.14 15,15	17	0.75	1.09	1,35E-001	0.90	0.10	9,44	SI	PMID: 15942958
chr5	SEC24A	134012148 17		GACGTGGCGGC 14.49 10.51 14,5	35	0.82	1.21	1,21E-002	0.90	0.10	9,34	NO	
chr18	VAPA	9903977 6 20 28		TACGTGCCGCC 17.49 12.87 17,5	44	0.08	0.91	7,96E-001	0.90	0.10	9,09	NO	
chr1	RORC	150065246 8 7 7		CACGTGGGGCC 14.06 11.45 18,05	11	0.43	0.87	4,46E-001	0.90	0.10	9,03	NO	
chr19	CHST8	388045248 5 9 8 6		CGCGTGGCGGC 12.53 15.02 15,67	8	0.83	1.58	5,30E-001	0.90	0.10	9,02	NO	
chr1	PTBP2	96959967 10 16		TGCGTGGCTCGG 10.37 16.84 16,85	24	0.01	0.74	9,79E-001	0.90	0.10	8,83	NO	
chr11	UVRAG	7,5E+07 9		CACGTGACGAG 17,95	19	0.17	0.84	6,85E-001	0.90	0.10	8,81	NO	

chr18	ANKRD12	9126708	7 12 17 22	GGCGTGACCGC	13.36 16.01 16.02	49	0.63	1.08	7.53E-003	0.90	0.10	8.81	NO	
chr11	MTCH2	4.8E+07	8	CACGTGACCTAG	17.25	29	0.05	0.62	8.48E-001	0.90	0.10	8.78	NO	
chr6	KLHDC3	4308692	6 7 14 7 7	GACGTGTACACC	13.72 14.87 16.84	31	0.01	0.61	9.79E-001	0.90	0.10	8.75	NO	
chr8	SLCO5A1	7.1E+07	13	CACGTGCCAAGC	17.12	10	0.09	0.69	8.74E-001	0.90	0.10	8.64	NO	
chrX	HUWE1	5360647	6 11	GACGTGCAGGG	16.99 11.39 17	45	0.04	0.93	9.11E-001	0.89	0.11	8.5	NO	
chr20	COX4I2	3E+07	8	TACGTGCACACA	17.98	13	0.23	0.81	6.37E-001	0.89	0.11	8.26	NO	
chr22	TEF	4009308	5 14 8 16 8	GGCGTGGAGCTT	8.61 10.76 17.24	29	0.09	0.92	8.23E-001	0.89	0.11	8.25	SI	PMID: 15073166
chr12	FOXN4	1.1E+08	13	CACGTGCCACAG	17.21	6	0.1	0.37	8.29E-001	0.89	0.11	8.21	NO	
chr5	IRF1	1318539	7 7	GGCGTGGACCG	11.35 17.13 17.13	18	0.11	1.02	8.49E-001	0.89	0.11	8.2	SI	PMID: 11325839
chr3	AMOTL2	1.4E+08	9	CACGTGCCGAAC	17.61	17	0.19	0.97	7.09E-001	0.89	0.11	8.08	NO	
chr8	PINX1	1.1E+07	8	TACGTGCCAAGG	17.44	13	0.16	0.79	7.55E-001	0.89	0.11	8.08	NO	
chr3	ST6GAL1	1881315	8	CACGTGCCACA	18.44 18.21 18.45	33	0.14	0.8	6.05E-001	0.89	0.11	8.05	NO	
chr7	EMID2	1007928	9 15 10	GGCGTGGGCTT	5.11 11.48 16.91	8	0.08	0.68	9.10E-001	0.89	0.11	7.95	NO	
chr9	SHB	3798520	3 12 8	GACGTGAGTGA	15.34 16.49 16.5	35	0.57	1.11	6.79E-002	0.89	0.11	7.95	NO	
chr7	PPP1R9A	9437503	7 21	CACGTGCCGTGC	16.75 12.41 16.76	23	0.02	0.81	9.57E-001	0.89	0.11	7.79	NO	
chr14	RCOR1	1E+08	9	TACGTGGTTGAM	12.48	24	1.09	0.88	6.27E-004	0.89	0.11	7.75	NO	
chr2	ZC3H6	1127497	3 8 13 7	GGCGTGCAGCG	15.35 12.85 17.46	15	0.2	0.97	7.24E-001	0.89	0.11	7.72	NO	
chr10	LDB1	1038646	5 14 5 7 16 2	TGCGTGTGTT	8.04 7.14 11.16 8.5	23	0.04	0.74	9.16E-001	0.88	0.12	7.62	NO	
chrX	TCEAL7	1024717	8 16	CACGTGAGGATT	13.58 14.95 14.96	6	0.99	0.97	3.57E-001	0.88	0.12	7.54	NO	
chr10	FAM107E1	1.5E+07	7	CGCGTGCAGAG	13.87	16	0.91	1	5.30E-002	0.88	0.12	7.39	NO	
chr9	KIAA2024	5997134	8	CACGTGCCCGAG	16.73	19	0.05	1.05	9.41E-001	0.88	0.12	7.27	NO	
chr3	SLC38A3	5021858	6 7 8 16 7 8	TGCGTGCCTTG	10.90 14.02 17.78	13	0.34	0.94	5.44E-001	0.88	0.12	7.19	NO	
chr3	THRB	2.4E+07	6	CACGTGCCGCT	16.91	22	0.09	1.05	8.68E-001	0.88	0.12	7.19	NO	
chr8	PEBP4	2266449	8 9 8 7 10	TGCGTGGCGTC	9.71 9.13 11.16 5.4	3	0.55	0.68	7.90E-001	0.87	0.13	6.97	NO	
chr19	SYT5	6E+07	4	CACGTGCAGGG	18.13	16	0.33	0.91	4.83E-001	0.87	0.13	6.93	NO	
chr10	EGR2	6424700	6 16 22 28	TGCGTGGGGCC	11.25 9.68 17.39	16	0.47	0.78	2.07E-001	0.87	0.13	6.93	NO	
chr22	FAM19A	4726398	6 5 7 8	CACGTGCCGGC	17.14 13.48 17.15	12	0.42	1.29	6.07E-001	0.87	0.13	6.88	NO	
chr1	GPR25	1991090	8 13 18	AGCGTGAGCAG	12.16 15.47 15.47	8	0.7	0.85	3.20E-001	0.87	0.13	6.76	NO	
chr11	BARX2	1287658	8 8	TACGTGCTATT	14.05 17.24 17.25	9	0.47	0.62	3.05E-001	0.87	0.13	6.72	NO	
chr3	ZIC4	1485886	10 21 15 6	GGCGTGAAGCA	11.74 8.11 16.66	12	0.06	0.93	9.31E-001	0.87	0.13	6.71	NO	
chr6	TFAP2B	5089739	11 11 10 12	AGCGTGAAGCAG	12.70 10.94 16.7	14	0.1	1.29	9.11E-001	0.87	0.13	6.61	NO	
chr19	FBL	4501669	8 9 18 5 8 16	CACGTGAGGGG	16.17 16.17 17.64	22	0.17	0.78	6.17E-001	0.87	0.13	6.56	NO	

IX.ANEXO V

Anexo V. Resultado de la clasificación funcional de los 216 genes del listado final usando el algoritmo de DAVID

Functional Group 1	Category	Term	Count	Geo: 1.189715879661195E-6	%	PValue	FDR	Genes
Functional Group 1	GOTERM_BP_ALL	GO:0006915~apoptosis	28	13.08%	6.29848E-07	0.001204883	0.001204883	SH2, TNFAIP3, AP1P1, PIM1, BCL2L11, PIM2, IGF1R, MNT, BCL6, FOXO3, MIF, RYBP, SHB, GAI1, VEGFA, DDIT4, BNIP3L, CXCR4, BRCC3, EGLN3, IGF1R3.
			28	13.08%	7.48487E-07	0.001431834	0.001431834	SH2, TNFAIP3, AP1P1, PIM1, BCL2L11, PIM2, IGF1R, MNT, BCL6, FOXO3, MIF, RYBP, SHB, GAI1, VEGFA, DDIT4, GADD45B, BTG1, BCAR1, CDKN1B, ACVR1B, BNIP3, JMY, BNIP3L, CXCR4, BRCC3, EGLN3, IGF1R3.
			28	13.08%	2.06147E-06	0.003949483	0.003949483	SH2, TNFAIP3, AP1P1, PIM1, BCL2L11, PIM2, IGF1R, MNT, BCL6, FOXO3, MIF, RYBP, SHB, GAI1, VEGFA, DDIT4, GADD45B, BTG1, BCAR1, CDKN1B, ACVR1B, BNIP3, JMY, BNIP3L, CXCR4, BRCC3, EGLN3, IGF1R3.
			28	13.08%	2.06147E-06	0.003949483	0.003949483	SH2, TNFAIP3, AP1P1, PIM1, BCL2L11, PIM2, IGF1R, MNT, BCL6, FOXO3, MIF, RYBP, SHB, GAI1, VEGFA, DDIT4, GADD45B, BTG1, BCAR1, CDKN1B, ACVR1B, BNIP3, JMY, BNIP3L, CXCR4, BRCC3, EGLN3, IGF1R3.
Functional Group 2	GOTERM_BP_ALL	GO:0006355~regulation of transcription, DNA-dependent	53	24.77%	1.06356E-05	0.020343824	0.020343824	CNLI, MXI1, BCKDHA, IRF1, RIF, PHF20, FOXO3, RYBP, RORA, NFIL3, HIVEP2, ZNF653, ZBTB25, KL4, KLF7, THRB, TCEAL7, FOXD1, CITED2, RORC, RCOR2, HOXA4, PHF21A, MNT, YEATS2, BCOR, BCL6, ZFX, TCEAL3, ZFY, ZNF292, TSC2D3, RCOR1, EGR2, TEF, RASSF1, HOXA11, KLF6, ATF4, BMP2, TFAP2B, SMAD7, FOXM4, JUNB, BARX2.
			53	24.77%	2.0085E-05	0.038415244	0.038415244	CNLI, MXI1, BCKDHA, IRF1, RIF, PHF20, FOXO3, RYBP, RORA, NFIL3, HIVEP2, ZNF653, ZBTB25, KL4, KLF7, THRB, LDB1, ID2, NR4A3, JMY, PNC1, MAFF, ELF3, KLF10, TCEAL7, FOXD1, CITED2, RORC, RCOR2, HOXA4, PHF21A, MNT, YEATS2, BCOR, BCL6, ZFX, TCEAL3, ZFY, ZNF292, TSC2D3, RCOR1, EGR2, TEF, RASSF1, HOXA11, KLF6, ATF4, BMP2, TFAP2B, SMAD7, FOXM4, JUNB, BARX2.
			53	24.77%	2.07789E-05	0.039738354	0.039738354	CNLI, MXI1, BCKDHA, IRF1, RIF, PHF20, FOXO3, RYBP, RORA, NFIL3, HIVEP2, ZNF653, ZBTB25, KL4, KLF7, THRB, LDB1, ID2, NR4A3, JMY, PNC1, MAFF, ELF3, KLF10, TCEAL7, FOXD1, CITED2, RORC, RCOR2, HOXA4, PHF21A, MNT, YEATS2, BCOR, BCL6, ZFX, TCEAL3, ZFY, ZNF292, TSC2D3, RCOR1, EGR2, TEF, RASSF1, HOXA11, KLF6, ATF4, BMP2, TFAP2B, SMAD7, FOXM4, JUNB, BARX2.
			53	24.77%	2.07789E-05	0.039738354	0.039738354	CNLI, MXI1, BCKDHA, IRF1, RIF, PHF20, FOXO3, RYBP, RORA, NFIL3, HIVEP2, ZNF653, ZBTB25, KL4, KLF7, THRB, LDB1, ID2, NR4A3, JMY, PNC1, MAFF, ELF3, KLF10, TCEAL7, FOXD1, CITED2, RORC, RCOR2, HOXA4, PHF21A, MNT, YEATS2, BCOR, BCL6, ZFX, TCEAL3, ZFY, ZNF292, TSC2D3, RCOR1, EGR2, TEF, RASSF1, HOXA11, KLF6, ATF4, BMP2, TFAP2B, SMAD7, FOXM4, JUNB, BARX2.
Functional Group 3	GOTERM_BP_ALL	GO:0019318~hexose metabolic process	12	5.61%	4.55786E-06	0.00871838	0.00871838	PKP, PKM2, HK1, GAPDH, PDK3, LDHA, ATF4, PKFB3, UGP2, PGAM1, PGK1, HK2.
			12	5.61%	6.07933E-06	0.011617926	0.011617926	PKP, PKM2, HK1, GAPDH, PDK3, LDHA, ATF4, PKFB3, UGP2, PGAM1, PGK1, HK2.
			12	5.61%	0.001686032	3.176530205	3.176530205	PKP, PKM2, HK1, GAPDH, PDK3, LDHA, ATF4, PKFB3, UGP2, PGAM1, PGK1, HK2.
			12	5.61%	0.001686032	3.176530205	3.176530205	PKP, PKM2, HK1, GAPDH, PDK3, LDHA, ATF4, PKFB3, UGP2, PGAM1, PGK1, HK2.
Functional Group 4	GOTERM_BP_ALL	GO:0046365~monosaccharide catabolic process	8	3.74%	8.59545E-06	0.016441693	0.016441693	PKP, PKM2, HK1, GAPDH, LDHA, PGAM1, PGK1, HK2.
			8	3.74%	2.32933E-05	0.044550418	0.044550418	PKP, PKM2, HK1, GAPDH, LDHA, PGAM1, PGK1, HK2.
			8	3.74%	5.88131E-05	0.112448537	0.112448537	PKP, PKM2, HK1, GAPDH, LDHA, PGAM1, PGK1, HK2.
			8	3.74%	6.35218E-05	0.124466307	0.124466307	PKP, PKM2, HK1, GAPDH, LDHA, PGAM1, PGK1, HK2.
Functional Group 5	GOTERM_BP_ALL	GO:004275~cellular carbohydrate catabolic process	8	3.74%	7.38654E-05	0.141208861	0.141208861	PKP, PKM2, HK1, GAPDH, LDHA, PGAM1, PGK1, HK2.
			8	3.74%	0.00034434	0.656665251	0.656665251	PKP, PKM2, HK1, GAPDH, LDHA, PGAM1, PGK1, HK2.
			8	3.74%	0.000500549	0.953210201	0.953210201	PKP, PKM2, HK1, GAPDH, LDHA, PGAM1, PGK1, HK2.
			8	3.74%	0.000500549	0.953210201	0.953210201	PKP, PKM2, HK1, GAPDH, LDHA, PGAM1, PGK1, HK2.
Functional Group 6	GOTERM_BP_ALL	GO:0051270~regulation of cell motility	5	2.34%	0.00970708	17.01881619	17.01881619	BCL6, EGR, BCAR1, CDKN1B, CXCR4.
			5	2.34%	0.01741624	20.22376849	20.22376849	BCL6, EGR, BCAR1, CDKN1B, CXCR4.
			5	2.34%	0.012290455	21.06704179	21.06704179	BCL6, EGR, BCAR1, CDKN1B, CXCR4.
			5	2.34%	0.012290455	21.06704179	21.06704179	BCL6, EGR, BCAR1, CDKN1B, CXCR4.
Functional Group 7	GOTERM_BP_ALL	GO:0045859~regulation of protein kinase activity	8	3.74%	0.02049013	32.48482696	32.48482696	CAV1, SPRY1, GADD45B, EGR, CDKN1B, CXCR4, AVP1, EDN1.
			8	3.74%	0.023089288	36.03752942	36.03752942	CAV1, SPRY1, GADD45B, EGR, CDKN1B, CXCR4, AVP1, EDN1.
			8	3.74%	0.025545712	39.04509843	39.04509843	CAV1, SPRY1, GADD45B, EGR, CDKN1B, CXCR4, AVP1, EDN1.
			8	3.74%	0.025545712	39.04509843	39.04509843	CAV1, SPRY1, GADD45B, EGR, CDKN1B, CXCR4, AVP1, EDN1.
Functional Group 7	GOTERM_BP_ALL	GO:0045893~positive regulation of transcription, DNA-dependent	9	4.21%	0.013640481	23.10554178	23.10554178	THRB, IRF1, BCKDHA, TEF, RASSF1, FOXO3, ATF4, CITED2, TFAP2B.
			9	4.21%	0.013640481	23.10554178	23.10554178	THRB, IRF1, BCKDHA, TEF, RASSF1, FOXO3, ATF4, CITED2, TFAP2B.

GOTERM_BP_ALL	GO:0045941~positive regulation of transcription	9	4.21%	0.041767382	55,78780543	THRB, IRF1, BCKDHA, TEF, RASSF1, FOXO3, ATF4, CITED2, TFAP2B,
GOTERM_BP_ALL	GO:0045935~positive regulation of nucleobase, nucleoside, nucleotide	9	4.21%	0.047965812	60,94958112	THRB, IRF1, BCKDHA, TEF, RASSF1, FOXO3, ATF4, CITED2, TFAP2B,
Functional Group 8						
Category		Geo: 0.03633886342030624				
GOTERM_BP_ALL	GO:0030097~hemopoiesis	7	3.27%	0.027797334	41,68375171	LDB1, BCL6, KLF6, ID2, FOXO3, TP52, ACVR1B,
GOTERM_BP_ALL	GO:0048534~hemopoietic or lymphoid organ development	7	3.27%	0.036988082	51,37314709	LDB1, BCL6, KLF6, ID2, FOXO3, TP52, ACVR1B,
GOTERM_BP_ALL	GO:0002520~immune system development	7	3.27%	0.046671209	59,9210662	LDB1, BCL6, KLF6, ID2, FOXO3, TP52, ACVR1B,
Functional Group 9						
Category		Geo: 0.04135752875914331				
GOTERM_BP_ALL	GO:0006469~negative regulation of protein kinase activity	4	1.87%	0.031824564	46,13541952	CAV1, SPRY1, GADD45B, CDKN1B,
GOTERM_BP_ALL	GO:0033673~negative regulation of kinase activity	4	1.87%	0.031824564	46,13541952	CAV1, SPRY1, GADD45B, CDKN1B,
GOTERM_BP_ALL	GO:0051348~negative regulation of transferase activity	4	1.87%	0.034695999	49,11067716	CAV1, SPRY1, GADD45B, CDKN1B,
GOTERM_BP_ALL	GO:0043086~negative regulation of catalytic activity	4	1.87%	0.083255703	81,04105824	CAV1, SPRY1, GADD45B, CDKN1B,
Functional Group 10						
Category		Geo: 0.06205232038883219				
GOTERM_BP_ALL	GO:0009250~glucan biosynthetic process	3	1.40%	0.013386832	22,72638601	PPP1R3C, GBE1, GYS1,
GOTERM_BP_ALL	GO:0005978~glycogen biosynthetic process	3	1.40%	0.013386832	22,72638601	PPP1R3C, GBE1, GYS1,
GOTERM_BP_ALL	GO:0033692~cellular polysaccharide biosynthetic process	3	1.40%	0.04644665	59,74008105	PPP1R3C, GBE1, GYS1,
GOTERM_BP_ALL	GO:000271~polysaccharide biosynthetic process	3	1.40%	0.04644665	59,74008105	PPP1R3C, GBE1, GYS1,
GOTERM_BP_ALL	GO:0005977~glycogen metabolic process	3	1.40%	0.07853045	79,08168606	PPP1R3C, GBE1, GYS1,
GOTERM_BP_ALL	GO:006073~glucan metabolic process	3	1.40%	0.082037514	80,55326981	PPP1R3C, GBE1, GYS1,
GOTERM_BP_ALL	GO:0043284~biopolymer biosynthetic process	3	1.40%	0.151827461	95,71536638	PPP1R3C, GBE1, GYS1,
GOTERM_BP_ALL	GO:0044264~cellular polysaccharide metabolic process	3	1.40%	0.185660881	98,03338738	PPP1R3C, GBE1, GYS1,
GOTERM_BP_ALL	GO:0005976~polysaccharide metabolic process	3	1.40%	0.194283754	98,39572751	PPP1R3C, GBE1, GYS1,
Functional Group 11						
Category		Geo: 0.06347310837202115				
GOTERM_BP_ALL	GO:0042325~regulation of phosphorylation	4	1.87%	0.058349579	68,33877494	PDGFB, EGFR, CDKN1B, IGFBP3,
GOTERM_BP_ALL	GO:0051174~regulation of phosphorus metabolic process	4	1.87%	0.06620119	73,02568461	PDGFB, EGFR, CDKN1B, IGFBP3,
GOTERM_BP_ALL	GO:0019220~regulation of phosphate metabolic process	4	1.87%	0.06620119	73,02568461	PDGFB, EGFR, CDKN1B, IGFBP3,
Functional Group 12						
Category		Geo: 0.06933827346208621				
GOTERM_BP_ALL	GO:0006171~cAMP biosynthetic process	3	1.40%	0.021133527	33,54294793	ADM, EDN1, ADCY5,
GOTERM_BP_ALL	GO:0046058~cAMP metabolic process	3	1.40%	0.030313569	44,5043245	ADM, EDN1, ADCY5,
GOTERM_BP_ALL	GO:009190~cyclic nucleotide biosynthetic process	3	1.40%	0.061765257	70,46592409	ADM, EDN1, ADCY5,
GOTERM_BP_ALL	GO:0009165~nucleotide biosynthetic process	3	1.40%	0.584169376	99,99999487	ADM, EDN1, ADCY5,
Functional Group 13						
Category		Geo: 0.08233145948432086				
GOTERM_BP_ALL	GO:0045860~positive regulation of protein kinase activity	5	2.34%	0.076322819	78,10187556	GADD45B, EGFR, CXCR4, AVP11, EDN1,
GOTERM_BP_ALL	GO:0033674~positive regulation of kinase activity	5	2.34%	0.082367578	80,92674529	GADD45B, EGFR, CXCR4, AVP11, EDN1,
GOTERM_BP_ALL	GO:0051347~positive regulation of transferase activity	5	2.34%	0.088132203	82,88003428	GADD45B, EGFR, CXCR4, AVP11, EDN1,
Functional Group 14						
Category		Geo: 0.1132027363856715				
GOTERM_BP_ALL	GO:0007584~response to nutrient	3	1.40%	0.08592007	81,94429978	BCKDHA, STC1, STC2,
GOTERM_BP_ALL	GO:0031667~response to nutrient levels	3	1.40%	0.151827461	95,71536638	BCKDHA, STC1, STC2,
GOTERM_BP_ALL	GO:0009991~response to extracellular stimulus	3	1.40%	0.177095823	97,59760587	BCKDHA, STC1, STC2,
Functional Group 15						
Category		Geo: 0.1421919143979914				
GOTERM_BP_ALL	GO:0001932~regulation of protein amino acid phosphorylation	3	1.40%	0.139486347	94,35159399	PDGFB, EGFR, IGFBP3,
GOTERM_BP_ALL	GO:0006521~regulation of amino acid metabolic process	3	1.40%	0.143575336	94,84351312	PDGFB, EGFR, IGFBP3,
GOTERM_BP_ALL	GO:0033238~regulation of amine metabolic process	3	1.40%	0.143575336	94,84351312	PDGFB, EGFR, IGFBP3,
Functional Group 16						
Category		Geo: 0.1724217988184875				

Category	Term	Count	%	PValue	FDR	Genes
GOTERM_BP_ALL	GO:0009187~cyclic nucleotide metabolic process	4	1.87%	0.016363835	27.06687235	ADM, RORA, EDNI, ADCY5,
GOTERM_BP_ALL	GO:0009117~nucleotide metabolic process	4	1.87%	0.542123909	99.99996763	ADM, RORA, EDNI, ADCY5,
GOTERM_BP_ALL	GO:0055085~nucleobase, nucleoside and nucleotide metabolic process	4	1.87%	0.577820397	99.99999315	ADM, RORA, EDNI, ADCY5,
Functional Group 17	Geo: 0.17638099462147755					
Category	Term	Count	%	PValue	FDR <td>Genes</td>	Genes
GOTERM_BP_ALL	GO:0045445~myoblast differentiation	3	1.40%	0.058578452	68.48665887	BTG1, IGFBP3, BAXX2,
GOTERM_BP_ALL	GO:0048747~muscle fiber development	3	1.40%	0.119460636	91.22893315	BTG1, IGFBP3, BAXX2,
GOTERM_BP_ALL	GO:0048741~skeletal muscle fiber development	3	1.40%	0.119460636	91.22893315	BTG1, IGFBP3, BAXX2,
GOTERM_BP_ALL	GO:0042692~muscle cell differentiation	3	1.40%	0.123405692	91.95091074	BTG1, IGFBP3, BAXX2,
GOTERM_BP_ALL	GO:0007519~skeletal muscle development	3	1.40%	0.242423968	99.50632411	BTG1, IGFBP3, BAXX2,
GOTERM_BP_ALL	GO:0014706~striated muscle development	3	1.40%	0.339239128	99.96390268	BTG1, IGFBP3, BAXX2,
GOTERM_BP_ALL	GO:0007517~muscle development	3	1.40%	0.625976866	99.99999932	BTG1, IGFBP3, BAXX2,
Functional Group 18	Geo: 0.21511941303980292					
Category	Term	Count	%	PValue	FDR <td>Genes</td>	Genes
GOTERM_BP_ALL	GO:0016568~chromatin modification	6	2.80%	0.092652885	84.43272055	BNIP3, RASSF1, BCOR, HUWE1, RCOR1, PHF21A,
GOTERM_BP_ALL	GO:0006325~establishment and/or maintenance of chromatin architecture	6	2.80%	0.320075214	99.93762626	BNIP3, RASSF1, BCOR, HUWE1, RCOR1, PHF21A,
GOTERM_BP_ALL	GO:0006323~DNA packaging	6	2.80%	0.335681821	99.95999794	BNIP3, RASSF1, BCOR, HUWE1, RCOR1, PHF21A,
Functional Group 19	Geo: 0.31831397954366114					
Category	Term	Count	%	PValue	FDR <td>Genes</td>	Genes
GOTERM_BP_ALL	GO:0055082~cellular chemical homeostasis	5	2.34%	0.251073582	99.60368555	BNIP3, STCI, CXCR4, EDNI, ABCB6,
GOTERM_BP_ALL	GO:0006873~cellular ion homeostasis	5	2.34%	0.251073582	99.60368555	BNIP3, STCI, CXCR4, EDNI, ABCB6,
GOTERM_BP_ALL	GO:0050801~ion homeostasis	5	2.34%	0.309527432	99.91626723	BNIP3, STCI, CXCR4, EDNI, ABCB6,
GOTERM_BP_ALL	GO:0019725~cellular homeostasis	5	2.34%	0.526164648	99.99993765	BNIP3, STCI, CXCR4, EDNI, ABCB6,
Functional Group 20	Geo: 0.32910591693112334					
Category	Term	Count	%	PValue	FDR <td>Genes</td>	Genes
GOTERM_BP_ALL	GO:0009615~response to virus	4	1.87%	0.114461932	90.22577356	BNIP3, BNIP3L, CXCR4, TBK1,
GOTERM_BP_ALL	GO:0051707~response to other organism	4	1.87%	0.430049071	99.95786594	BNIP3, BNIP3L, CXCR4, TBK1,
GOTERM_BP_ALL	GO:0009607~response to biotic stimulus	4	1.87%	0.724149086	100	BNIP3, BNIP3L, CXCR4, TBK1,
Functional Group 21	Geo: 0.33159284177839893					
Category	Term	Count	%	PValue	FDR <td>Genes</td>	Genes
GOTERM_BP_ALL	GO:0030005~cellular di-, tri-valent inorganic cation homeostasis	4	1.87%	0.297669558	99.88402565	STCI, CXCR4, EDNI, ABCB6,
GOTERM_BP_ALL	GO:0055066~di-, tri-valent inorganic cation homeostasis	4	1.87%	0.300843376	99.89365105	STCI, CXCR4, EDNI, ABCB6,
GOTERM_BP_ALL	GO:0030003~cellular cation homeostasis	4	1.87%	0.365752723	99.98350938	STCI, CXCR4, EDNI, ABCB6,
GOTERM_BP_ALL	GO:0055080~cation homeostasis	4	1.87%	0.369111981	99.98510236	STCI, CXCR4, EDNI, ABCB6,
Functional Group 22	Geo: 0.337340481737406					
Category	Term	Count	%	PValue	FDR <td>Genes</td>	Genes
GOTERM_BP_ALL	GO:0031214~biomineral formation	3	1.40%	0.326167637	99.94749229	THRB, EGFR, BMP2,
GOTERM_BP_ALL	GO:0001503~ossification	3	1.40%	0.326167637	99.94749229	THRB, EGFR, BMP2,
GOTERM_BP_ALL	GO:0046849~bone remodeling	3	1.40%	0.360847457	99.98089074	THRB, EGFR, BMP2,
Functional Group 23	Geo: 0.44170125731275095					
Category	Term	Count	%	PValue	FDR <td>Genes</td>	Genes
GOTERM_BP_ALL	GO:006874~cellular calcium ion homeostasis	3	1.40%	0.423927416	99.99738203	STCI, CXCR4, EDNI,
GOTERM_BP_ALL	GO:0055074~calcium ion homeostasis	3	1.40%	0.423927416	99.99738203	STCI, CXCR4, EDNI,
GOTERM_BP_ALL	GO:0006875~cellular metal ion homeostasis	3	1.40%	0.460220296	99.99924605	STCI, CXCR4, EDNI,
GOTERM_BP_ALL	GO:0055065~metal ion homeostasis	3	1.40%	0.460220296	99.99924605	STCI, CXCR4, EDNI,
Functional Group 24	Geo: 0.5328569232348714					
Category	Term	Count	%	PValue	FDR <td>Genes</td>	Genes
GOTERM_BP_ALL	GO:0030030~cell projection organization and biogenesis	4	1.87%	0.532856923	99.99995251	GJA1, PDGFB, CXCR4, DSCAM1L,
GOTERM_BP_ALL	GO:0048858~cell projection morphogenesis	4	1.87%	0.532856923	99.99995251	GJA1, PDGFB, CXCR4, DSCAM1L,
GOTERM_BP_ALL	GO:0032990~cell part morphogenesis	4	1.87%	0.532856923	99.99995251	GJA1, PDGFB, CXCR4, DSCAM1L,

Functional Group 25	Category	Term	Geo: 0.576049614446131	Count	%	PValue	FDR	Genes
	GOTERM_BP_ALL	GO:0048667~neuron morphogenesis during differentiation	GJA1, CXCR4, DSCAML1,	3	3	1.40%	0.524834958	99,99993422 GJA1, CXCR4, DSCAML1,
	GOTERM_BP_ALL	GO:0048812~neurite morphogenesis	GJA1, CXCR4, DSCAML1,	3	3	1.40%	0.524834958	99,99993422 GJA1, CXCR4, DSCAML1,
	GOTERM_BP_ALL	GO:0031175~neurite development	GJA1, CXCR4, DSCAML1,	3	3	1.40%	0.594261473	99,9999968 GJA1, CXCR4, DSCAML1,
	GOTERM_BP_ALL	GO:0048666~neuron development	GJA1, CXCR4, DSCAML1,	3	3	1.40%	0.672692332	99,9999995 GJA1, CXCR4, DSCAML1,



## IX. ANEXO VI

**Anexo VI. SNPs que se localizan en sitios RCGTG.** SNP\_ID, es el número identificativo del SNP. Validación, el proyecto en el que ha sido validado. Alelo, cambio de base.SNP\_HBS, posición del cambio en el HBS.Max HBS, si el SNP coincide con el HBS de máxima puntuación.

SNP_ID	Validación	Alelo	SNP_HBS	Gen	Chr	HBS_Loc	Max_HBS
rs13358075	hapmap	T/A	ACG_G	SPOCK1	5	136374995	NO
rs2901215	NA	T/G	ACG_G	TIAL1	10	121346556	NO
rs56306258	NA	A/T	_CGTG	ANKRD12	18	9126754	NO
rs56033752	NA	G/A	AC_TG	DLG2	11	83576587	NO
rs16943318	cluster,freq,hapmap	G/A	CAC_C	RORA	15	58997259	NO
rs3758554	freq	C/G	CA_GC	LDB1	10	103864659	NO
rs34394782	NA	C/A/T	CACG_	CITED2	6	139737161	NO
rs34476700	NA	G/-	_CGTG	SENP3	17	7404132	NO
rs56298217	NA	G/A	GC_TG	CLK3	15	72701699	NO
rs17152486	freq,hapmap	C/T	A_GTG	FLJ23834	7	105459125	SI
rs58433430	NA	C/T	A_GTG	KLHL14	18	28519311	SI
rs17004038	cluster,freq	C/A	_ACGT	MIF	22	22566590	SI
rs35867052	NA	C/T	A_GTG	KCNMA1	10	78431889	NO
rs41305185	NA	G/A	AC_TG	ARX	X	24944039	NO
rs8028749	cluster,freq,doub	C/T	A_GTG	NPTN	15	71671759	NO
rs17151058	freq,hapmap	G/A	AC_TG	UPF2	10	12003288	NO
rs16958416	freq,hapmap	G/A	AC_TG	STX8	17	9364007	NO
rs3867457	NA	G/A	AC_TG	CTBP2	10	126667223	NO
rs4776356	hapmap	A/C	_CGTG	IQCH	15	65474427	NO
rs10450768	NA	C/T	A_GTG	CORO1C	12	107564638	NO
rs55687074	NA	G/A	AC_TG	CDH13	16	81849926	NO
rs62588589	NA	A/T	_CGTG	GK	X	30622217	NO
rs210963	NA	A/G	_CGTG	MYB	6	135545387	NO
rs3762460	cluster,freq,doub	A/G	_CGTG	ZNF593	1	26368135	NO
rs7123260	NA	C/G	A_GTG	PAX6	11	31775041	NO
rs2799759	freq,hapmap	G/A	AC_TG	TBX15	1	119249382	NO
rs17084199	freq,hapmap	C/A	CA_GC	MCTP1	5	94245707	NO
rs45472197	NA	C/T	CA_GC	TJP1	15	27780131	NO
rs10482607	NA	C/G	CA_GC	NR3C1	5	142763379	NO
rs4381672	cluster,freq,doub	A/G	C_CGC	ZNF521	18	20966787	NO
rs8111884	cluster,doublehit	C/T	CA_GC	ZNF536	19	35551882	NO
rs12491302	cluster,hapmap	G/A	CAC_C	ZBTB20	3	115945606	NO
rs62377608	NA	G/A	CAC_C	PPP2R2B	5	146415810	NO
rs10024529	cluster,freq,hapmap	G/A	CAC_C	KIAA0746	4	25472785	NO
rs13138842	cluster	A/G	C_CGC	GRIA2	4	158362543	NO
rs12859504	NA	C/T	CACG_	FGF13	X	137622084	NO
rs41298547	freq	C/G	CACG_	PGRMC2	4	129428759	NO
rs1804995	NA	C/A	CA_GC	WDR1	4	9685176	NO
rs62191421	NA	C/A	CACG_	SOX12	20	253196	NO
rs10428011	NA	G/A	CAC_C	CELSR1	22	45309712	NO
rs10421122	cluster,freq,doub	C/A	CACG_	MAST1	19	12807323	NO
rs62108604	NA	C/T	CA_GC	CACNA1A	19	13239072	NO
rs1529745	freq,doublehit	C/G	CACG_	GATAD2A	19	19378320	NO
rs3862027	cluster,freq,hapmap	C/T	CA_GC	PAX2	10	102520171	NO
rs13430871	NA	C/T	CACG_	MAP4K4	2	101680362	NO
rs34908810	NA	C/A	_ACGC	CYP1B1	2	38157735	NO
rs61882397	NA	C/G	_ACGC	KCNC1	11	17759942	NO
rs7036453	NA	C/T	CA_GC	BNC2	9	16722732	NO
rs17126220	freq,hapmap	C/T	CA_GC	ROR1	1	64375245	NO

rs7951227	doublehit,hapma	C/G	CA_GC	PLEKHA7	11	16765660	NO
rs2014408	cluster,freq,hapn	C/T	CA_GC	SOX6	11	16321855	NO
rs45538937	NA	G/A	CAC_T	NPAS3	14	32914447	NO
rs12497972	hapmap	C/T	CA_GT	EPHB1	3	136253280	NO
rs5993939	NA	C/T	CA_GT	RTN4R	22	18608959	NO
rs10408467	NA	C/T	_ACGT	WIZ	19	15394041	NO
rs816845	cluster,doublehit	C/T	CA_GT	KCNMA1	10	78948880	NO
rs61945842	NA	G/A	CAC_T	PXN	12	119177283	NO
rs9790917	freq,hapmap	G/A	CAC_T	SLC1A3	5	36665667	NO
rs6702094	cluster,freq,doub	G/A	CAC_T	DNM3	1	170558152	NO
rs11542156	NA	G/A	CAC_T	UBE2Q1	1	152789421	NO
rs1380432	NA	C/T	G_GTG	SMOC1	14	69524278	NO
rs61732122	NA	G/A	GC_TG	SORCS3	10	106392370	NO
rs3763763	cluster,freq,doub	C/A	G_GTG	TACC2	10	124003592	NO
rs10489045	cluster,freq,hapn	G/A	GCGT_	AFF1	4	88213806	NO
rs11134527	cluster,freq,doub	G/A	GC_TG	SLIT3	5	168127931	NO
rs1803142	NA	C/T	G_GTG	ARF1	1	226353191	NO
rs11589148	NA	G/T	_CGTG	CASZ1	1	10779499	NO
rs9341238	NA	G/C	_CGTG	CYP1B1	2	38157629	NO
rs1786315	NA	G/A	GC_TG	PABPC1	8	101804028	NO
rs12799093	NA	G/A	GC_TG	KIRREL3	11	125841244	NO
rs2756892	cluster,doublehit	C/T	G_GTG	TPM2	9	35680308	NO
rs13198379	NA	C/G	G_GTG	TFEB	6	41810542	NO
rs690170	cluster,doublehit	G/C	AC_TG	MAP1A	15	41597711	SI
rs10050309	freq,hapmap	C/T	A_GTG	AFF1	4	88154897	SI
rs640249	cluster,freq,subn	G/T	ACGT_	ACTC1	15	32878973	SI
rs17544103	cluster,freq,hapn	C/T	A_GTG	SLIT2	4	20228986	SI
rs4885499	cluster,freq,doub	G/A	AC_TG	EDNRB	13	77441719	SI
rs1140830	NA	T/A	ACG_G	EPB41L3	18	5383186	SI
rs13311220	NA	G/A	AC_TG	PUS7	7	104949920	SI
rs17329324	cluster,freq,hapn	C/G	A_GTG	SEMA5A	5	9600956	SI
rs2487700	hapmap	G/A	AC_TG	DLEU7	13	50185235	SI
rs11623672	cluster	C/G	A_GTG	TTC7B	14	90351386	SI
rs12962319	NA	G/T	AC_TG	NEDD4L	18	54074450	SI
rs2810008	cluster,freq,doub	G/A	AC_TG	PLD5	1	240754010	SI
rs12089784	cluster,freq	G/A	AC_TG	GPX7	1	52841015	SI
rs12224101	NA	C/T	A_GTG	MAML2	11	95385205	SI
rs479759	hapmap	G/A	AC_TG	KCNC4	1	110577374	SI
rs1049851	NA	G/C	CAC_C	HNRNPA2B	7	26198045	SI
rs17154780	hapmap	A/T	C_CGC	PCOLCE	7	100037885	SI
rs6101647	freq,hapmap	C/A	CA_GC	NDRG3	20	34715003	SI
rs34183373	NA	A/-	C_CGC	PANK1	10	91333892	SI
rs45454797	NA	C/G	_ACGC	CHMP2A	19	63758202	SI
rs7253731	NA	C/A	CA_GC	AKT2	19	45482908	SI
rs13051078	NA	C/T	CACG_	ABCG1	21	42512216	SI
rs60155532	NA	G/A	CAC_C	ILDR1	3	123204826	SI
rs11811223	NA	C/A	CACG_	SHC1	1	153213399	SI
rs1024924	cluster,freq,doub	G/A	CAC_C	FAM49A	2	16701977	SI
rs61747198	NA	G/A	CAC_C	KIAA1958	9	114420029	SI
rs34743335	cluster	A/T	C_CGC	IRF6	1	208046150	SI

rs59684124	NA	C/T	CA_GC	TNS1	2	218514530	SI
rs41270959	NA	G/A	CAC_C	IPO9	1	200070800	SI
rs2666099	NA	A/G	C_CGC	DES	2	219997481	SI
rs4832508	cluster,freq,doub	G/T	CAC_C	KCNS3	2	17922627	SI
rs35867246	NA	C/-	CACG_	ZFP91	11	58103790	SI
rs17860583	NA	C/T	CA_GC	MMP13	11	102320042	SI
rs6492946	cluster,freq,doub	G/A	CAC_T	BAHD1	15	38519399	SI
rs3788865	NA	G/A	CAC_T	NDP	X	43706989	SI
rs4799349	NA	A/T	C_CGT	ASXL3	18	29412558	SI
rs11552199	NA	C/T	CA_GT	MTMR12	5	32263487	SI
rs34370865	freq	C/T	CA_GT	PTEN	10	89614718	SI
rs1062751	NA	C/T	CA_GT	MORF4L1	15	76977019	SI
rs12134821	freq,hapmap	G/A	CAC_T	DNM3	1	170634159	SI
rs10917503	cluster,doublehit	C/G	_ACGT	NBL1	1	19838356	SI
rs809621	NA	C/T	_ACGT	MFHAS1	8	8788574	SI
rs13290547	cluster,freq,hapn	C/T	CA_GT	DAB2IP	9	123527313	SI
rs6694643	cluster,freq,hapn	A/T	C_CGT	C8A	1	57121993	SI
rs8139784	freq,hapmap	G/T	GCGT_	RANGAP1	22	40011955	SI
rs12578436	cluster,doublehit	G/A	GC_TG	ACVRL1	12	50591964	SI
rs62369145	NA	G/T	GCGT_	HCN1	5	45732347	SI
rs1648309	NA	G/A	GCGT_	SHF	15	43246469	SI
rs34595128	freq	C/T	G_GTG	SOCS1	16	11257140	SI
rs41305365	NA	G/A	_CGTG	PCTK1	X	46963046	SI
rs62379870	NA	G/A	_CGTG	GALNT10	5	153662167	SI
rs3138162	cluster,freq,hapn	C/A	G_GTG	DCN	12	90097473	SI
rs1803426	NA	G/A	GCGT_	TM9SF4	20	30216960	SI
rs1902598	hapmap	C/T	G_GTG	DMXL2	15	49702479	SI
rs6864748	cluster,doublehit	G/A	GC_TG	ERGIC1	5	172299593	SI
rs7232067	NA	C/T	G_GTG	SERPINB5	18	59298291	SI
rs11552890	NA	G/C	GCGT_	KIFAP3	1	168310248	SI
rs12528746	NA	G/T	_CGTG	COL11A2	6	33267380	SI
rs6679980	cluster,freq,doub	C/T	G_GTG	RCOR3	1	209500293	SI
rs59549221	NA	G/A	GC_TG	DIXDC1	11	111359228	SI
rs12528746	NA	G/T	_CGTG	COL11A2	6	33267380	SI
rs12528746	NA	G/T	_CGTG	COL11A2	6	33267380	SI
rs12528746	NA	G/T	_CGTG	COL11A2	6	33267380	SI
rs12528746	NA	G/T	_CGTG	COL11A2	6	33267380	SI
rs12528746	NA	G/T	_CGTG	COL11A2	6	33267380	SI
rs10624	NA	G/T	GCGT_	CDC2L6	6	111038119	SI
rs2934623	NA	G/T	_CGTG	CUL3	2	225157844	SI
rs11573545	NA	C/T	G_GTG	PTCH2	1	45077979	SI
rs35712197	NA	-/C	_ACGT	NA	5	NA	NA
rs11438611	NA	-/T	_ACGT	NA	X	NA	NA
rs35853371	NA	-/G	_ACGT	NA	18	NA	NA
rs7943121	cluster,freq,doub	G/C	ACGT_	NA	11	NA	NA
rs34233798	NA	-/T	_GCGT	NA	12	NA	NA
rs2763140	hapmap	G/C	_CGTG	NA	6	NA	NA



***ARTÍCULOS RELACIONADOS***



# Genome-wide identification of hypoxia-inducible factor binding sites and target genes by a probabilistic model integrating transcription-profiling data and *in silico* binding site prediction

Amaya Ortiz-Barahona<sup>1</sup>, Diego Villar<sup>1</sup>, Nuria Pescador<sup>1</sup>, Jorge Amigo<sup>2</sup> and Luis del Peso<sup>1,\*</sup>

<sup>1</sup>Department of Biochemistry, Universidad Autónoma de Madrid-Instituto de Investigaciones Biomédicas CSIC, Madrid and <sup>2</sup>Spanish National Genotyping Center (CeGen), Genomic Medicine Group, CIBERER, Universidad de Santiago de Compostela, Galicia, Spain

Received November 4, 2009; Revised December 10, 2009; Accepted December 11, 2009

## ABSTRACT

The transcriptional response driven by Hypoxia-inducible factor (HIF) is central to the adaptation to oxygen restriction. Hence, the complete identification of HIF targets is essential for understanding the cellular responses to hypoxia. Herein we describe a computational strategy based on the combination of phylogenetic footprinting and transcription profiling meta-analysis for the identification of HIF-target genes. Comparison of the resulting candidates with published HIF1a genome-wide chromatin immunoprecipitation indicates a high sensitivity (78%) and specificity (97.8%). To validate our strategy, we performed HIF1a chromatin immunoprecipitation on a set of putative targets. Our results confirm the robustness of the computational strategy in predicting HIF-binding sites and reveal several novel HIF targets, including RE1-silencing transcription factor co-repressor (RCOR2). In addition, mapping of described polymorphisms to the predicted HIF-binding sites identified several single-nucleotide polymorphisms (SNPs) that could alter HIF binding. As a proof of principle, we demonstrate that SNP rs17004038, mapping to a functional hypoxia response element in the macrophage migration inhibitory factor (MIF) locus, prevents induction of this gene by hypoxia. Altogether, our

results show that the proposed strategy is a powerful tool for the identification of HIF direct targets that expands our knowledge of the cellular adaptation to hypoxia and provides cues on the inter-individual variation in this response.

## INTRODUCTION

Cells respond to chronic hypoxia by altering their gene expression pattern to optimize metabolic oxygen consumption, maintain energy balance and restore oxygen supply. Many of the genes involved in this adaptive response are directly regulated by the hypoxia-inducible factor (HIF) (1), a transcription factor that is activated when oxygen tension drops. HIF is a heterodimer composed of an oxygen-regulated alpha subunit (HIF $\alpha$ ) (2) and a constitutively expressed beta subunit (HIF $\beta$ , also known as Aryl receptor nuclear translocator, ARNT) (3) that partners with a number of basic-helix-loop-helix transcription factors. Oxygen affects both HIF $\alpha$  half-life (4) and transactivation (5). In normoxia, HIF $\alpha$  is hydroxylated at two proline residues (6,7) by a family of dioxygenases (EGL nine homologs, EGLNs) that require oxygen as cosubstrate (8,9). This posttranslational modification labels HIF $\alpha$  for proteosomal degradation, as the proline-hydroxylated form is recognized by an E3-ubiquitin ligase complex that contains the VHL tumor suppressor (10). In addition, another dioxygenase (factor inhibiting HIF, FIH) catalyzes the oxygen-dependent hydroxylation of an asparagine residue, located in the C-terminal transactivation domain, preventing its

\*To whom correspondence should be addressed. Tel: +34 91 585 4440; Fax: +34 91 585 4400; Email: luis.peso@uam.es; lpeso@iib.uam.es

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

© The Author(s) 2009. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.5>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

interaction with the p300 coactivator and blunting HIF $\alpha$  transcriptional activity (11–13). In hypoxia, all these hydroxylation reactions become compromised, due to the reduced availability of oxygen, resulting in HIF $\alpha$  stabilization and recruitment of coactivators, such as p300. Thus, under hypoxia, HIF accumulation allows its interaction with HIF $\beta$  and its binding to the RCGTG motif, known as hypoxia response element (HRE), within regulatory regions of its target genes. There are three genes encoding for HIF $\alpha$  subunits: HIF1 $\alpha$ , HIF2 $\alpha$  (also known as EPAS) and HIF3 $\alpha$ . HIF1 $\alpha$  and HIF2 $\alpha$  have been extensively studied, while HIF3 $\alpha$  remains poorly characterized. The regulation of HIF1 $\alpha$  and 2 $\alpha$  by hypoxia is similar and both bind to the same core motif (1). However, recent evidence indicates that these transcription factors induce overlapping but not identical sets of genes (14,15), suggesting nonredundant functions for HIF1 $\alpha$  and HIF2 $\alpha$ .

Given the central role of HIF in the transcriptional response to hypoxia, the characterization of HIF target genes provides critical insights into the adaptations required to cope with reduced oxygen tension. Over a hundred HIF-targets have been described (1) as the result of research efforts focused on individual genes. These studies revealed that many of the genes regulated by hypoxia are involved in the reprogramming of cellular metabolism and restoration of oxygen supply. More recently, a number of studies described the effect of hypoxia in the transcriptome by means of gene expression profiling. These studies, covering a wide range of cell types and conditions (16–26), revealed a large number of novel potential targets. Although undoubtedly relevant, a major drawback of this approach is that it cannot distinguish between direct and secondary HIF targets. In addition, no attempts have been made to combine the results of all these studies. Such integrative studies, or meta-analysis, have higher statistical power to detect relevant effects than single studies and provide a generalization to the individual experiments. In fact, several works (27) have demonstrated that the application of meta-analysis to multiple independent gene expression data sets leads to the identification of sets of significant, differentially expressed genes, void of the artifacts of individual studies. Finally, two recent reports (28,29) coupled transcript profiling and chromatin immunoprecipitation (ChIP) followed by hybridization to genomic tiling microarrays (ChIP–Chip) to identify direct HIF targets. A comparative analysis is needed to reveal the extent of overlap between conclusions of both studies and also whether further studies are required. Thus, in spite of intense research efforts, the complete characterization of HIF targets is still unresolved.

*In silico* identification of transcription-factor-binding sites (TFBS) is a powerful tool to complement experimental identification of transcription factor targets (30). These methods rely on the comparison of candidate sequences to a position-specific scoring matrix (PSSM) constructed by alignment of known binding sites for the transcription factor of interest. HIF binds to a short, but extremely well-conserved [A/G]CGTG motif. Conservation of other positions outside this motif is controversial: while some studies suggest that some positions show a base

distribution significantly different from random expectation (1,31,32), other studies fail to find conservation outside the core RCGTG (28,29). Nevertheless, the low information content of most PSSMs, including that of HIF, and the size of mammalian genomes result in a large number of potential hits across the genome. Since conserved noncoding sequences (CNS) are enriched in *cis*-regulatory elements (33,34), a successful approach to reduce the number of spurious hits is to restrict the search for TFBS to these regions. The identification of CNSs, based on multiple species alignment of noncoding genomic sequences, reveals evolutionarily conserved regions (phylogenetic footprinting) that may have been selected during evolution due their regulatory or structural function. The algorithm PhastCons implements a two state hidden Markov model that provides a score value that reflects the conservation of each base of a reference genome within a multiple species alignment (35). Therefore, potential regulatory regions can be inferred from PhastCons elements (groups of adjacent nucleotides with a significant conservation score).

Recent works (36–38) have demonstrated that the combination of gene expression data and TFBS prediction is a powerful tool for the identification of transcription factor target genes. In the present study, we applied a probabilistic model that integrates the evidence for the regulation of each particular gene by hypoxia (transcript profiling meta-analysis) and the presence of high-scoring HIF-binding sites (HBSs) for the identification of novel HIF targets. The application of this strategy results in a list of 216 predicted targets, most of them not previously reported as regulated by hypoxia. We tested the accuracy of our strategy by experimentally validating several of the identified HBSs by ChIP–quantitative polymerase chain reaction (qPCR). Moreover, we demonstrated that RCOR2, one of the borderline targets identified, is indeed a HIF-target gene. In addition, the strategy reported herein provides the coordinates for several hundred potential HBSs. We propose that, in addition to the identification of HIF-target genes, this information can be useful to identify genome variants within the population that could have an altered hypoxic response. As a proof of principle, we found that one of these variants has a major impact on the hypoxic induction of macrophage migration inhibitory factor (MIF). Given the relevance of hypoxia in pathologies, such as cancer and cardiovascular disease, an altered response to hypoxia could be among the underlying causes explaining different clinical courses and/or response to treatments.

## MATERIALS AND METHODS

### Cell culture and hypoxic conditions

The cell lines HeLa, HepaC1/4 and HepG2 were maintained in Dulbecco's modified Eagle medium, while HepaC1 and HepaC4 cell lines were grown in MEM- $\alpha$  medium. In all cases, the culture medium was supplemented with 100 U/ml penicillin, 100  $\mu$ g/ml streptomycin and 10% (v/v) fetal bovine serum. Cells were grown at 37°C in a humidified atmosphere containing 5% CO<sub>2</sub>.



For hypoxia treatments, cells were grown at the indicated oxygen concentration in a Whitley hypoxystation (don Whitley Scientific, UK).

### Plasmid construction

Human genomic DNA extracted from HeLa cells was used as template for PCR amplification of MIF and RCOR2 promoter regions using primers 1 + 2 and 7 + 8 (Supplementary Table S1), respectively. Reporter constructs were generated by cloning the PCR products into the pGL3-Basic plasmid (Invitrogen). The identity of all constructs was verified by sequencing. The mutant HRE and single-nucleotide polymorphism (SNP) constructs were generated by site-directed mutagenesis, employing PCR QuikChange Site-direct mutagenesis kit (Stratagene). Primers harboring the desired mutation were 2 + 3 (HREmut MIF), 9 + 10 (HREmut RCOR2) and 5 + 6 (SNP-HREmut MIF), respectively (Supplementary Table S1).

### Reporter assays

Cells were plated in six-well plates 24 h prior transfection. Each plate was transfected with a DNA mixture containing 0.9  $\mu$ g (HeLa cells) or 1.9  $\mu$ g (HepG2 cells) of the indicated reporter plasmid and 0.1  $\mu$ g of a plasmid encoding the *Renilla* firefly luciferase under the control of a SV40 promoter. 12–13 h after transfection, cells were replated in 24-well plates and then transferred to hypoxic conditions (1% oxygen) or left under normoxic conditions for 24 h. Subsequently, firefly and renilla luciferase activities were determined using a dual luciferase system (Promega, Madison, WI, USA). In order to correct for transfection efficiency, the luciferase activity was normalized to the *Renilla* luciferase activity. Each experimental condition was assayed in duplicate.

### ChIP assays

For ChIP assays, HeLa cells were grown on 10-cm plates until they reached 85% confluence, at which point they were exposed to hypoxia (1% oxygen) or left under normoxic conditions for 6 h. Subsequently, cells were fixed for 12 min at 4°C by adding formaldehyde to culture media to final concentration of 1% (v/v). Cross-linking was stopped by the addition of glycine (0.125 M final). The cells were washed with cold phosphate-buffered saline (PBS) and then lysed by scraping in 1 ml of lysis buffer [1% sodium dodecyl sulfate (SDS), 10 mM EDTA, 50 mM Tris/HCl, pH 8.1 and a protease inhibitor cocktail, Roche]. Cell lysates were incubated on ice for 10 min and then sonicated to shear the DNA to fragments between 200 and 1500 bp. After the removal of the insoluble material by centrifugation, 50  $\mu$ l of each sample was removed and stored (input), while 100  $\mu$ l were diluted in 1-ml immunoprecipitation buffer (1% Triton X-100, 2 mM EDTA, 150 mM NaCl and 20 mM Tris/HCl, pH 8.1). The lysates were precleared with 200  $\mu$ g of a Salmon Sperm DNA/Protein A agarose 50% slurry (Upstate Biotechnology, Lake Placid, NY, USA) for 1 h at 4°C. The samples were then immunoprecipitated twice, initially

with whole rabbit serum for 6 h [immunoglobulin G (IgG) control] and then overnight at 4°C with a polyclonal anti-HIF1 alpha antiserum (Abcam, ab2185). Immunocomplexes were recovered by the addition of 400  $\mu$ g of Salmon Sperm DNA/Protein A agarose 50% slurry to the samples that were then sequentially washed for 15 min in TSE I (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris/HCl, pH 8.1 and 150 mM NaCl), TSE II (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris/HCl, pH 8.1 and 500 mM NaCl) and buffer III (0.25 M LiCl, 1% NP-40, 1% deoxycholate, 1 mM EDTA and 10 mM Tris/HCl, pH 8.1). Finally, the complexes were washed twice with TE buffer (10 mM Tris, pH 8.0 and 1 mM EDTA) and extracted twice with a buffer containing 1% SDS and 0.1 M NaHCO<sub>3</sub>. The eluates were pooled, and cross-linking was reversed by the addition of 200 mM NaCl (final concentration) and overnight incubation at 65°C. The proteins were removed by the addition of proteinase K (30  $\mu$ g/sample) for 2 h at 42°C, and the DNA was purified by phenol-chloroform extraction and ethanol precipitation. Immunoprecipitated DNA was amplified by qPCR using the primers (11–48) indicated in Supplementary Table S1.

### RNA extraction and qPCR

Cells were harvested in 1 ml of Ultraspec reagent (Biotecx). Subsequently RNA was reverse-transcribed to cDNA (Improm-II reverse transcriptase; Promega).

q-PCR was performed with the LC FastStart DNA master SYBR GreenI kit (Roche Applied Science) and in a Light Cycler system (Roche Applied Science) using the indicated primers (Supplementary Table SI). Data were analyzed with Light Cycler software version 3.5.28 (Idaho Technology). For each sample, duplicate determinations were made and the gene copy number was normalized to the amount of  $\beta$ -actin.

### Meta-analysis of gene profiling data sets

For the meta-analysis, we downloaded 16 independent experiments from Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) (39) database (Supplementary Table S2). For those experiments analyzing more than two conditions (for example the effect of hypoxia and HIF overexpression), we generated independent data sets for each comparison. Thus, we generated 19 data sets from the 16 experiments (Supplementary Table S2). In all the cases, untreated normoxic cells were used as reference. For each data set, we calculated the mean for each probe values in the biological replicates. Probes with null values were discarded. Then, for each probe, we calculated the effect of treatments (hypoxia, hypoxia mimetics or HIF expression) as the logarithm of the ratio of the means of treated and control samples. Finally, individual log-ratios values were normalized by subtraction of the mean of all the log-ratios across the data set and division by their standard deviation.

For the meta-analysis, each gene locus was treated independently and tested for the null hypothesis that no

gene was modulated by treatments. To this end, the normalized log-ratio values of all probes (across all data sets) mapping to the gene locus under consideration were compared to zero using one-sample *t*-test. The resulting *P*-values were corrected for multiple testing by applying false discovery rate. Genes with adjusted *P*-values <0.01 were considered significantly regulated by treatments. Custom Perl scripts were used for the analysis, complemented with R-based extensions for the statistical calculations.

### Identification and scoring of HBSs

For the identification of HBSs, we restricted our search to noncoding regions of genes. To this end, we considered all the RefSeq transcripts encoded by each locus and identified the intronic and untranslated regions within by projecting all transcripts, therefore excluding transcript-specific coding sequences. In addition, for each locus, we selected a 5-kb region upstream the transcription start site, TSS. When necessary, the upstream region was trimmed to avoid overlap with adjacent loci. For genes with several TSS, we selected the outermost TSS to define the 5-kb upstream region (the region upstream the remaining TSS is considered as part of the intronic regions). After the localization of all noncoding regions, we identified mammal or vertebrate PhastCons elements (35) within. Adjacent PhastCons elements were fused if more than 50% of the sequence in the resulting fused region was conserved. We refer to these PhastCons elements located in noncoding regions as conserved noncoding sequences (33) or CNSs. Then, we identified conserved RCGTG motifs within these CNSs. A motif was considered conserved when it was present at least in four mammals, including human and mouse. Sequences lacking conserved RCGTG motifs were discarded as potential HBSs. Finally, sequences containing a conserved motif were scored according to a PSSM. For the generation of this matrix, we selected 23 well-characterized HIF-binding sequences corresponding to 22 HIF-target genes, together with an orthologous sequence (Supplementary Table S4). Then, we used a chi-squared test to determine those positions with an observed distribution of residues significantly different to that expected by chance. This analysis revealed that, in addition to the RCGTG motif, some positions from -1 to +17 (being the R residue of the core motif the position +1) showed a significantly skewed distribution ( $P < 0.01$ ). For each position, we calculated the log-odds ratio of the observed frequencies of each nucleotide over the background frequency found for that nucleotide. The background frequency was obtained from the counts of each nucleotide in the CNSs: A, 0.275; C, 0.223; G, 0.229; T, 0.273. The log-odds ratios were arranged in a  $4 \times 18$  matrix (a column per position and a row for each nucleotide) so that the score for the nucleotide *i* at position *j* is:

$$S_{ij} = \log_2 \left( \frac{\text{freq}_{i,j}^{\text{observed}}}{\text{freq}_i^{\text{background}}} \right)$$

To calculate the score for the whole sequence (*S*) we added the individual scores for each position. Since not all positions had the same information content (Supplementary Table S4), the contribution of each position to the final score was weighted by the information content of the position (*I<sub>j</sub>*):

$$S = \sum_j I_j * S_{i,j}$$

The information content for each position *j* was calculated from the Shannon entropy:

$$I_j = - \sum \text{freq}_i^{\text{background}} * \log_2 \left( \text{freq}_i^{\text{background}} \right) - \sum \text{freq}_{i,j}^{\text{observed}} * \log_2 \left( \text{freq}_{i,j}^{\text{observed}} \right)$$

The RefSeq coordinates, PhastCons coordinates and the alignments corresponding to the identified CNSs were downloaded from the UCSC genomic browser (<http://genome.ucsc.edu/index.html>) (40,41). All coordinates correspond to the hg18 human genome assembly. The analysis was performed with custom scripts written in Perl.

### Classification of genes as HIF target/nontarget

To classify any given gene as a HIF target or nontarget, we calculated the relative likelihood that the gene belongs to any of these two groups. To this end, we constructed models that, given the fold induction of the gene and associated *P*-value according to our meta-analysis and the score of the potential HBSs found within the gene locus, assign a probability to the gene in each of the two states. Then, the relative likelihood of being a HIF target is the ratio (odds ratio) between the probabilities according to each model. In the HIF-target model (*T*), we fitted the distribution of fold induction ( $f_i^f(x)$ ) and HBS score ( $f_i^s(y)$ ) values for the set of well-characterized HIF-target genes (Supplementary Table S4) to a normal (Gauss) density function. For the nontarget (Background, *B*) model, we assumed that most of the genes in the genome are not regulated by HIF, thus calculated the Gaussian density functions describing the distribution of fold induction and HBS score values for all the analyzed genes [ $f_b^f(x)$  and  $f_b^s(y)$ , respectively]. Then, the probability of a gene being a HIF target given its fold induction (and *P*-value) and HBS score,  $P(x,y,p|T)$ , is the product of functions describing HBS score and fold induction:

$$P(x,y,p|T) = f_i^s(\max(y_i)) * \left( f_i^f(x) + \left( p * f_b^f(x) \right) \right)$$

where *x* and *p* are the meta-analysis values for the fold induction and associated *P*-value respectively and  $\max(y_i)$  is the maximum score of the HBSs found within the gene locus. Similarly the probability of the gene being a nontarget (background) is:

$$P(x,y,p|B) = f_b^s(\max(y_i)) * \left( f_b^f(x) + \left( p * f_i^f(x) \right) \right)$$

Finally, the ratio of these two likelihoods (odds ratio) represent the relative probability of being a HIF-target gene:

$$\frac{P(x,y,p|T)}{P(x,y,p|B)}$$

for simplicity we refer to this ratio as  $P_T/P_B$  ratio.

In these expressions, the contribution of the fold induction value ( $f^f(x)$ ) to the probability was weighted by the  $P$ -value associated to the mean so that when the fold induction is not reliable (for large  $P$ -values approaching 1) its contribution to the probability is very similar for target and nontarget:

$$f_i^f(x) + (p * f_b^f(x)) \approx f_b^f(x) + (p * f_i^f(x))$$

In this case, the classification (odds ratio) is just based on the score value:

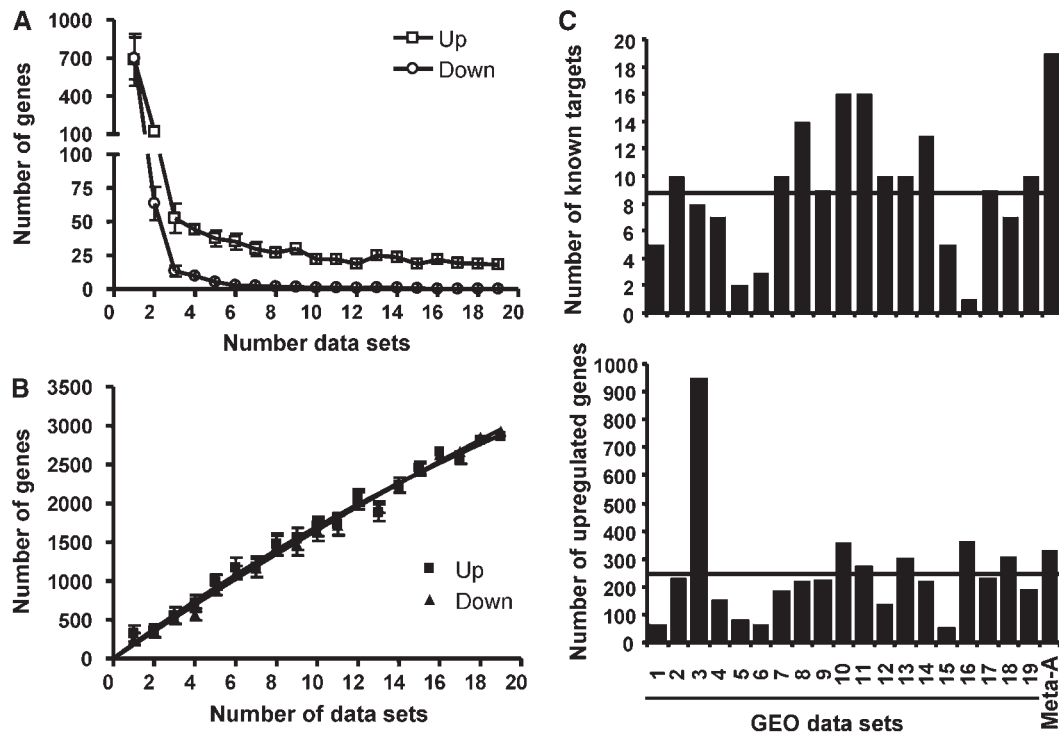
$$\frac{P(x,y,p|T)}{P(x,y,p|B)} \approx \frac{f_i^s(\max(y_i))}{f_b^s(\max(y_i))}$$

## RESULTS

### Meta-analysis of gene expression profile data sets from cells exposed to hypoxia

In order to identify HBSs, we designed a strategy based on the intersection of two independent approaches:

(i) identification of hypoxia-modulated genes through the analysis of multiple gene expression data sets from publicly available databases (transcription profiling meta-analysis); (ii) identification of evolutionarily conserved HIF-binding motifs within potential *cis*-regulatory regions (phylogentic footprinting). For the first approach, we selected 19 data sets from the GEO database (<http://www.ncbi.nlm.nih.gov/geo/>) (39) corresponding to 16 independent experiments that analyzed the gene expression profile of cells exposed to hypoxia or hypoxia mimetics (for simplicity, we refer to them as hypoxia herein). We processed each of these data sets to calculate the  $\log_2$  of the hypoxia/normoxia ratio (log-ratio) for each probe and then considered as significantly regulated by hypoxia those probes whose log ratio was  $>1.96$  or  $>2.6$  SD above or below the data set mean. A gene was considered modulated when at least one of its probes was significantly up- or downregulated. This analysis revealed that only a small group of genes was induced by hypoxia across all the experiments (Figure 1A and Supplementary Table S3). In addition, even when a relatively relaxed criterion ( $>1.96$  SD from the mean) was used to ascribe genes to the downregulated group, no gene was found consistently repressed by hypoxia in all the experiments (Figure 1A). On the other hand, the number of nonredundant genes modulated by hypoxia increased rapidly with the number of experiments (Figure 1B). When taking into consideration all 19 data sets, we



**Figure 1.** Comparison of individual gene profiling studies versus meta-analysis. The indicated number of data sets (number of data sets) was randomly selected out from the 19 GEO tables without replacement. The number of genes whose expression was 1.96 SD from the mean in all (A) or at least in one (B) of the selected data set was recorded and the procedure repeated 10 times. The graph represents the mean number of recorded genes and error bars the standard deviation. (C) For each individual data set (1 to 19, see Supplementary Table S2), the genes showing a fold induction ratio  $>2.6$  SD above the mean were considered upregulated. In the case of the meta-analysis (Meta-A), genes with a corrected  $P$ -value  $<0.01$  and mean fold induction positive were considered upregulated. The graph represents the number of known target genes (according to ref. 1) represented in the upregulated group in each case (upper graph) together with the total number of upregulated genes (lower graph). The horizontal lines in each graph represent the average number of known and upregulated genes across the 19 data sets.

found a total number of 2864 up- and 2929 downregulated nonredundant genes that together account for 49.6% of the genes represented in these data sets. These results indicate that the simple intersection of results from individual experiments is too restrictive, while their combination results in excessive noise, highlighting the need for a statistical analysis of the combined data sets. To this end, we treated each gene as an independent hypothesis, compared to the null hypothesis that the gene is not modulated by hypoxia and thus the mean value of the log-ratios of all its probes is 0. For each gene, we obtained the value of the log-ratio for all associated probes across all the data sets, calculated their mean (mean fold induction) and used one-sample *t*-test to ask whether it differed significantly from 0. After correction of the resulting *P*-values for multiple testing (false discovery rate), we selected genes with a *P*-value below 0.01 as regulated by hypoxia. This analysis resulted in a total of 259 (2.22%) genes induced and 191 (1.64%) genes repressed by hypoxia out of 11 673 genes represented in all GEO data sets. As a crude measure of the meta-analysis performance, we looked for known HIF targets (1) in the set of upregulated genes identified in each independent study or in our meta-analysis. As shown in Figure 1C, the meta-analysis excelled the performance of the individual studies, recovering a higher number of known targets than any of them. In addition, the increased sensitivity did not seem to be accompanied by a reduction of specificity, since the total number of upregulated genes identified by the meta-analysis was not different to the average number identified in individual studies.

### *In-silico* identification of HBSs

For the prediction of genome-wide HIF binding positions, we searched for the occurrence of RCGTG motifs in the human genome. Since CNSs are genomic regions enriched in *cis*-regulatory elements, we restricted our search to these regions in order to increase the chances of finding relevant motifs and to reduce the number of spurious hits. In addition, we only considered RCGTG motifs that were conserved in, at least, four species, including mouse. For each locus, we defined CNSs as PhastCons elements mapping to introns, untranslated regions and promoter regions upstream of the TSS. This search resulted in 9458 potential HBSs (conserved RCGTG motifs) distributed across 3980 genetic loci (34.1% of the analyzed genes). We found no conserved HBSs in the remaining 7693 (65.9%) gene loci analyzed.

Integration of the meta-analysis results and the presence of conserved HBSs showed that, as expected, the proportion of genes upregulated by hypoxia that contained at least one conserved HBS was significantly higher than expected by chance ( $P = 3.2 \times 10^{-14}$ , Table 1). In contrast, we found no significant association between the presence of an HBS and downregulation of the gene by hypoxia ( $P = 0.42$ , Table 1). Since we found no evidence for a direct role of HIF on downregulation of gene expression, we focused on genes upregulated by hypoxia throughout the study.

**Table 1.** The presence of an HBS correlates with transcriptional upregulation but not with repression by hypoxia

	HBS+		HBS-		<i>P</i> -value
	Obsr.	Expc.	Obsr.	Expc.	
Upregulated	145	88	113	170	$3.18 \times 10^{-14}$
Nonregulated	3766	3823	7459	7402	
Downregulated	69	64	121	126	0.42
Nonregulated	3766	3771	7459	7454	

Genes were classified as upregulated [mean  $\log_2(\text{hypoxia/normoxia}) > 0$  and  $P < 0.01$ ], downregulated [mean  $\log_2(\text{hypoxia/normoxia}) < 0$  and  $P < 0.01$ ] or nonregulated ( $P > 0.01$ ).

The number of genes in each group with at least one potential HBS (HBS+) or none (HBS-) is shown (Obsr.).

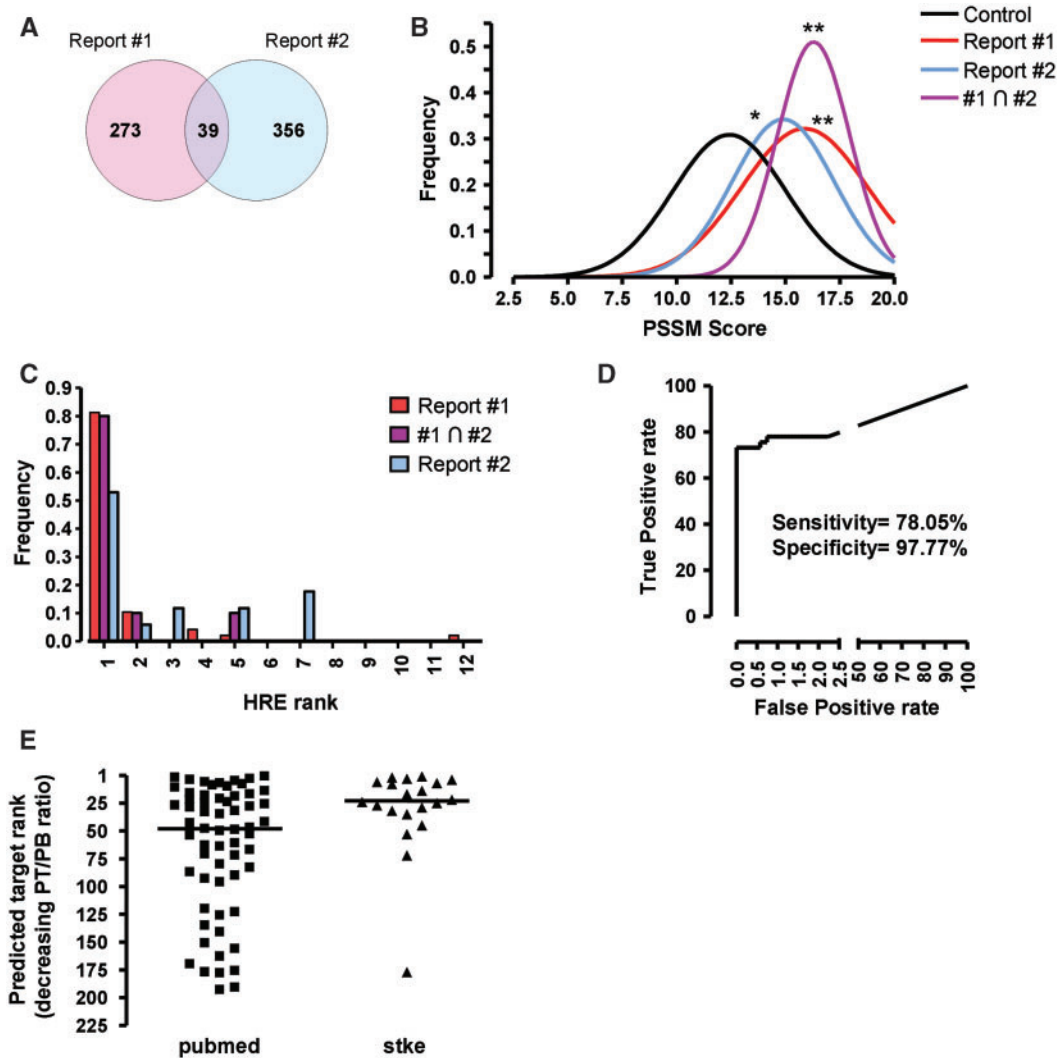
The number of genes expected by chance in each group is also shown (Expc.).

The significance of the difference between the observed versus expected frequencies was calculated by a Chi-squared test and the resulting *P*-value is shown (*P*-value).

### Scoring of HBSs

The alignment of a set of well-characterized HREs reveals that, in addition to the core RCGTG motif, other positions present a distribution of bases significantly different to that expected by chance (Supplementary Table S4 and ref. 31). Thus, we decided to use this information to infer functional HBSs. An 18-residue-long PSSM was generated based on the alignment of 46 sequences (see Supplementary Table S4 and 'Materials and Methods' section for details), and subsequently used it to assign a score value to each of the identified HBSs. In order to assess the ability of this score to discriminate functional HBSs, we studied the distribution of scores for HBSs recently identified by genome-wide ChIP-Chip (28,29). A comparative analysis reveals that there is very little overlap among the HIF-binding regions reported in these two studies (Figure 2A), probably because of the different cell lines/experimental conditions, data analysis and array platforms used in each work. Thus, we only considered the overlapping group of genes (Supplementary Table S5) as a reliable set of HIF targets. Figure 2B shows the distribution of scores of our predicted HBSs that map to any of the published HIF-binding regions (28,29). For comparison, we also plotted the score distribution for all the HBSs identified within CNSs across the genome (control). These results show that the score distribution for experimentally determined HBSs is shifted toward higher score values. In particular, the scores of the HBSs mapping to the regions identified in both reports (highly reliable HBSs) are higher than those of control genes, and the mean score for this group is significantly different from that of the controls (Figure 2B). It is worth pointing out that only four (GAPDH, LDHA, PGK1 and TF) of the 39 regions common to both ChIP-chip studies are coincident with the HRE regions used to construct the scoring matrix. Thus, the results shown in Figure 2B are not due to overfitting of our PSSM matrix to a specific set of HBSs.

Our search often predicted several HBSs for a given locus (Supplementary Table S6). Therefore, in order to



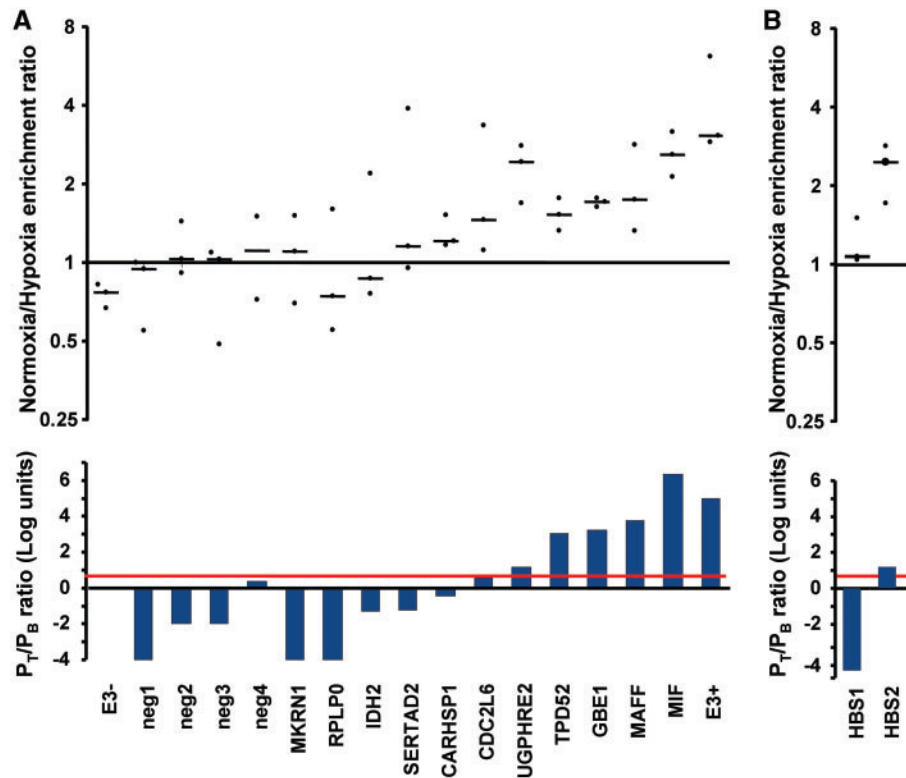
**Figure 2.** High HBS scores correlate with functional HIF-binding sites. (A) Venn diagram showing the number of overlapping HIF-binding sites identified by ChIP-chip in two published reports (ref. 29, Report #1; ref. 28, Report #2). (B) The scores of HBSs identified by our strategy were discretized (binning size 0.5 U) and their frequency distribution was calculated and adjusted to a Gauss curve by nonlinear fitting. The graph shows the resulting curves for all the HBSs identified across the genome (control), the HBSs mapping to HIF-binding regions identified by ChIP-chip in each report (Report #1, Report #2) or those HBS in regions common to both reports ( $\#1 \cap \#2$ ). The scores in each group were compared (ANOVA) and statistically significant differences with the control group are indicated by asterisks (\*,  $P < 0.01$ ; \*\*,  $P < 0.001$ ). (C) The potential HBSs identified for each gene were ranked according to their score in decreasing order (rank 1 corresponds to the highest scoring HBS) and the rank of the predicted HBSs mapping to HIF-binding sites was recorded. The figure shows the rank frequency distribution for predicted HBSs mapping to HIF-binding regions identified by ChIP-Chip in each report (Report #1, Report #2) or regions common to both reports ( $\#1 \cap \#2$ ). (D) Receiver operating characteristic (ROC) curve of known positive/negative (see text) targets versus prediction using a  $P_T/P_B$  ratio of 6.5 as threshold to classify genes as potential targets. (E) Genes identified as potential targets ( $P_T/P_B$  ratio  $> 6.5$ ) were sorted in decreasing  $P_T/P_B$  ratio order. The graph represents the rank of known HIF targets, according to ref. 1 (Stke) or a bibliographic search (PubMed), within the predicted target list. Horizontal line represents the median of each group.

further test the relevance of the score value, we next studied which one of the HBSs identified for each locus mapped to the sites bound by HIF according to ChIP-chip studies (28,29). To this end, the HBSs identified for each gene were ranked according to their score (rank 1 corresponded to the HBS with the highest score for a given locus) and the number of predicted HBSs of each rank that were coincident with an experimentally determined HBS was represented (Figure 2C). The data show that, in most cases, the experimentally validated HBS for a given locus matched the predicted HBS of highest score

value. Altogether, the results in Figure 2B and C support that the HBS score is a good predictor of functionality.

#### Probabilistic model of integrated binding site and gene expression data

When combined, the two approaches described above resulted in a list (Supplementary Table S6) in which each gene had associated parameters reflecting the magnitude of its modulation by hypoxia (fold induction and associated  $P$ -value) and one or several potential HBSs



**Figure 3.** Experimental validation of HIF binding to predicted sites. HeLa cells were exposed to normoxia or hypoxia (1% oxygen) for 6 h. After treatments, cells were processed for chromatin immunoprecipitation using antibodies directed to HIF1 $\alpha$ . The binding of HIF1 $\alpha$  to the predicted HBS within the indicated genes (A) was determined by qPCR. In the case of UGP2, HIF binding to two conserved HBSs was tested (B). The graph shows the ratio of the immunoprecipitated material in hypoxia over normoxia. The results from three independent experiments (black circles) and their median (line) are shown. In order to normalize data from the three independent experiments, the hypoxia/normoxia ratio is represented as fold over the mean value obtained for all the negative controls in each experiment. Neg1, IRS4; neg 2, STT3A; neg 3, HIVEP; neg4, LTBP1. The binding of HIF1 $\alpha$  to the HRE within EGLN3 enhancer (E3+) or to a nonfunctional RCGTG within EGLN3 locus (E3-) were used as internal controls (ref. 31). For comparison,  $P_T/P_B$  ratio (in logarithmic scale) for each target is shown (bottom histogram), along with the threshold value of 6.5 (red line).

mapping to regulatory regions within the locus, each of them having an associated score value. Our goal was to use this information to calculate a measure of the relative likelihood that the gene is an HIF target, as opposed of being nondirectly regulated by HIF (background). To this end, we constructed models that assign a probability to the gene in each of the two cases, and obtained the ratio of the two probabilities (odds ratio), we refer to this odds ratio as  $P_T/P_B$  ratio. In order to determine the optimum value for the  $P_T/P_B$  ratio for maximum sensitivity and specificity of gene classification, we used receiver operating characteristic (ROC) curve analysis (Figure 2D). For this analysis, the common set of 39 HIF targets from ChIP-Chip studies (Figure 2A and Supplementary Table S5) was used as known true targets. On the other hand, we selected genes that, while presenting conserved HBSs, were not induced by hypoxia (>30 probes in all data sets, mean fold induction between -0.3 and +0.3 and  $P > 0.5$ ) as negative set. According to this analysis, a  $P_T/P_B$  ratio >6.5 resulted in an optimum sensitivity of 78.05% and a selectivity of 97.77%. Thus, we calculated the  $P_T/P_B$  ratio for all genes represented in GEO data sets and classified them as HIF targets ( $P_T/P_B$  ratio >6.5) or nontargets ( $P_T/P_B$  ratio  $\leq 6.5$  or lack of HBSs). Through this strategy,

we predicted 216 HIF-target genes (Supplementary Table S6). Among them, 20 were previously known as HIF targets (1) and for 44 additional genes, some bibliographic evidence for their regulation by hypoxia was found (Supplementary Table S6). The remaining 152 genes are, to our knowledge, novel potential targets. The representation of the position of known target genes in our ranked list of predicted targets (Figure 2E) shows that they cluster toward the top positions [median values of 23 and 48, for the known targets from (1) and PubMed, respectively]. Thus, the  $P_T/P_B$  ratio accurately represents the probability of being an HIF target.

#### Experimental validation of model predictions

In order to evaluate the accuracy of our predictions we exposed HeLa cells to normoxia or hypoxia (1% oxygen) for 12 h and determined HIF1 $\alpha$  binding to a set of predicted HBSs by ChIP-qPCR. For this purpose, we randomly selected six genes (RPLP0, MAFF, IDH2, SERTAD2, TPD52 and CARHSP1) among those that, according to our meta-analysis, were significantly upregulated by hypoxia ( $P < 0.01$ ). To simplify the validation, we restricted our selection to genes having a single potential HBS. In addition, we included in this analysis the HBSs identified for the UDP-glucose

Pyrophosphorylase (UGP2) gene due to its potential role in glycogen metabolism (manuscript submitted for publication) and the HREs identified within MIF and CDC2L6 genes because of their inter-individual variation (see subsequent discussion). Finally, we also included five HBS motifs located in genes that were not induced by hypoxia in any of the GEO data sets (IRS4, STT3A, HIVEP1 and LTBP1) as a negative control group to estimate the background hypoxic/normoxic enrichment ratio. After the treatments, cells were processed for ChIP using an antibody directed to HIF1 $\alpha$ . Then, we determined the quantity of each of the target sequences in the HIF1 $\alpha$  immunoprecipitate by qPCR and expressed it as percentage of the input of immunoprecipitated chromatin. Finally, we calculated the ratio for the enrichment in samples exposed to hypoxia over normoxic samples. As expected, the hypoxic/normoxic enrichment ratio for the negative controls was close to 1 (Figure 3A). In addition, we found a high enrichment ratio for the HBS in the EGLN3 enhancer (31), used as positive control. Importantly, in spite of the high variability observed for the biological replicates, we found a good correlation between target prediction (Figure 3A bottom histogram) and experimental determination of HIF binding (Figure 3A, top graph). In general, only the candidates above the threshold (UGP2, TPD52, GBE1, MAFF1 and MIF) showed a consistent positive enrichment ratio in all three independent experiments. In contrast, the genes predicted as negative (MKRN1, RPLP0, IDH2 and SERTAD2) showed a pattern similar to that of negative controls. There were two exceptions, CARHSP1 and CDC2L2, that did not behave as predicted. It is important to note, however, that in both cases the  $P_T/P_B$  ratio was close to the threshold value of 6.5.

To further test the strength of our classifier, we determined HIF binding to the two potential HBSs identified within the UGP2 gene, HBS\_1 and HBS\_2, located in chromosome 2 at positions 63 922 445 and 63 923 300, respectively. According to their  $P_T/P_B$  ratios,  $3 \times 10^{-5}$  and 14.4, respectively, only one of them (HBS\_2) was classified as an HBS. In agreement with our prediction, only HBS\_2, but not HBS\_1, was consistently found in HIF1 $\alpha$  immunoprecipitates (Figure 3B). Collectively, these results confirm the accuracy of our predictions.

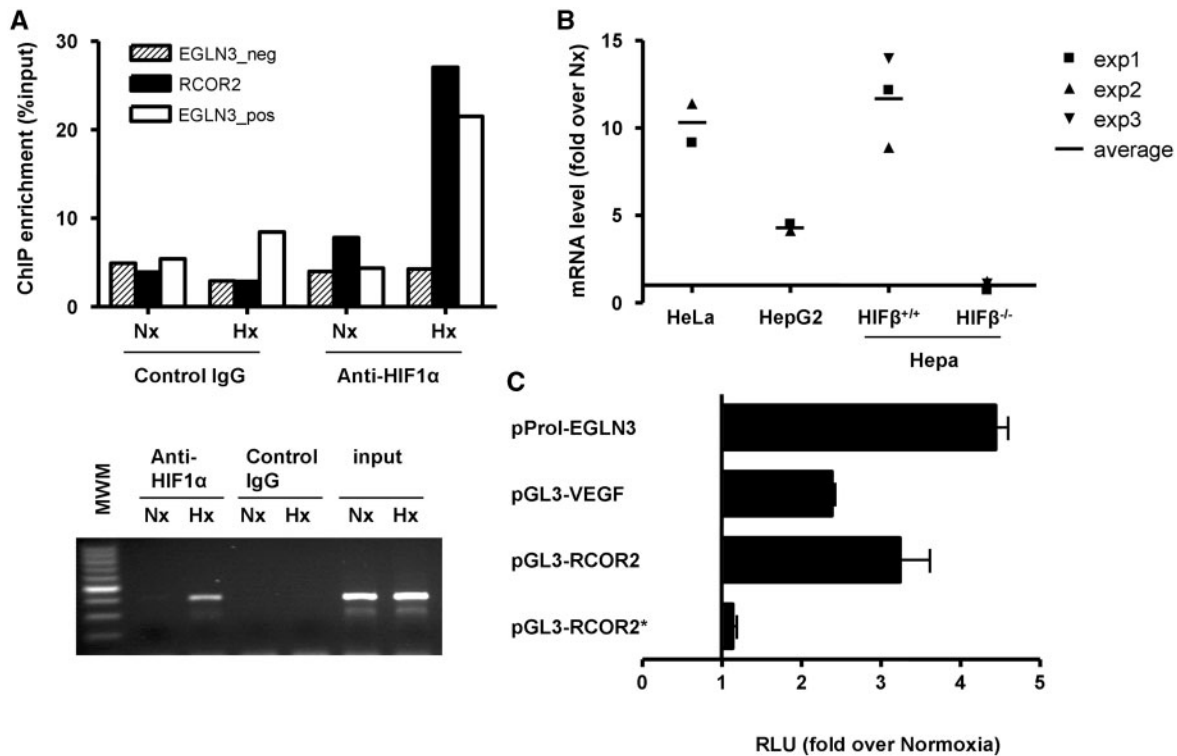
Next, we decided to determine the biological relevance of the identified HBSs. To this end, we focused on RE1-silencing transcription factor co-repressor 2 (RCOR2) because it was classified as positive by our strategy with a  $P_T/P_B$  ratio close to the threshold (9.63). In addition, it was not found significantly upregulated by hypoxia in the gene expression profiling meta-analysis (fold induction = 0.44 and  $P = 0.399$ ). Thus, the classification of RCOR2 as a true target was mainly based on the score value of its HBS (see 'Materials and Methods' section).

First, we investigated HIF1 $\alpha$  binding to the potential HBS by ChIP-qPCR. As shown in Figure 4A, chromatin from the target region was enriched in HIF1 $\alpha$ -immunoprecipitates from cells exposed to hypoxia. The enrichment was similar to that observed for the EGLN3 enhancer and was not observed when a control IgG was used for the immunoprecipitation (Figure 4A).

To determine whether HIF1 $\alpha$  binding to this site had a functional effect, we measured levels of RCOR2 mRNA in cells exposed to hypoxia by qPCR. Figure 4B shows that RCOR2 mRNA was induced by hypoxia in several cell lines. Moreover, the induction of RCOR2 was dependent on functional HIF since it was observed in HepaC1 cells, but not in HepaC1 derivative lacking HIF $\beta$  (42). Finally, we generated a reporter construct (pGL3-RCOR2) by cloning the RCOR2 promoter region, containing the putative HRE, upstream a firefly luciferase gene. Transfection of this construct into HepG2 cells demonstrated that the promoter activity was induced by hypoxia (Figure 4C). The induction was of similar magnitude to that observed for other HIF-regulated regions such as VEGFA promoter and EGLN3 enhancer (Figure 4C). Importantly, the mutation of the putative HRE completely abolishes the regulation of RCOR2 promoter by hypoxia (Figure 4C). Thus, RCOR2 is a novel hypoxia regulated gene whose induction under low oxygen is dependent on HIF activity and the presence of the HRE identified by our computational strategy. In addition, these results further support the relevance of our HIF-target predictions and show the robustness of our approach even for borderline cases.

#### Identification of SNPs that interfere with the response to hypoxia

The adaptation to hypoxia is largely dependent on HIF-mediated gene expression. Therefore, the identification of SNPs mapping to HBSs could reveal individuals with an altered response to hypoxia. The strategy described above provided a catalog of genome-wide HBSs (Supplementary Table S6). Thus, we decided to use this information to search for SNPs mapping to these sites. We retrieved from the dbSNP (<http://www.ncbi.nlm.nih.gov/projects/SNP/>) entries whose genomic coordinates were coincident with the RCGTG motifs identified by our computational strategy. This search resulted in 146 SNP mapping to HBS motifs (Supplementary Table S7). Among them, 12 corresponded to sites within potential HIF targets (Table 2). We focused on rs17004038 and rs10624 because they were validated SNPs and were located within the only HRE identified for MIF and CDC2L6 genes, respectively. As shown in Figure 3, both sites bound HIF *in vivo*; thus, rs17004038 and rs10624 map to functional HBSs. In order to investigate the biological effect of these polymorphisms, we cloned the wild type MIF promoter region or the C $\rightarrow$ A variant, corresponding to the SNP rs17004038, upstream a firefly luciferase gene and performed reporter assays with these constructs. As previously reported (43), MIF promoter region (WT) was robustly induced by hypoxia (Figure 5). Mutation of the HBS identified by our strategy completely abrogated luciferase induction (Figure 5, mutHRE), demonstrating its role in the transcriptional upregulation of MIF promoter and in agreement with its binding to HIF (Figure 3). Importantly, the variant allele C $\rightarrow$ A was not upregulated by hypoxia and its behavior was indistinguishable from the mutant HRE (Figure 5). In contrast to the strong effect on the regulation by hypoxia,



**Figure 4.** RCOR2 is a HIF-target gene. (A) HeLa cells were exposed to normoxia (Nx) or hypoxia (Hx, 1% oxygen) for 12h. After treatments, cells were processed for chromatin immunoprecipitation using antibodies directed to HIF1α (anti-HIF1α) or control immunoglobulins (control IgG). The binding of HIF1α to the predicted HRE within RCOR2, to the HRE within EGLN3 enhancer (EGLN3\_pos) or to a nonfunctional RCGTG within EGLN3 locus (EGLN3\_neg) were determined by quantitative (upper panel) and semi-quantitative PCR (RCOR2, lower panel). MWM, molecular weight marker. (B) HeLa, HepG2 and Hepa C1/C4 cells were exposed to normoxia or hypoxia for 12h and the level of RCOR2 mRNA was determined by quantitative PCR. The amount of each mRNA in samples was normalized to the content of β-actin mRNA in the same sample. The graph represents the fold values of hypoxic over normoxic mRNA levels normalized to the value of 1 (horizontal axis). Data represents the values from three independent experiments and their average (horizontal bar). (C) HepG2 cells were transfected with a reporter plasmid containing RCOR2 promoter region (−1770 to −795) upstream a luciferase reporter gene. Where indicated (asterisk) the consensus HRE sequence (ACGT) was mutated to TAGC. For comparison, reporter constructs containing the EGLN3 enhancer and VEGF promoter were included. The graphs represent the corrected luciferase activity values of each hypoxic sample over the luciferase activity obtained in normoxic cells. Data shown are a representative experiment out of three independent determinations.

**Table 2.** SNP mapping to RCGTG motifs within potential HIF-targets

SNP_ID	Validation	Allele	SNP_HBS	Gene	Chr	Coordin	Max_HBS
rs17004038	cluster, freq	C/A	_ACGT	MIF	22	24 236 591	YES
rs17152486	freq, hapmap	C/T	A_GTG	FLJ23834	7	105 671 891	YES
rs3758554	freq	C/G	CA_GC	LDB1	10	103 874 672	NO
rs16943318	cluster, freq, hapmap	G/A	CAC_C	RORA	15	61 209 971	NO
rs13358075	hapmap	T/A	ACG_G	SPOCK1	5	136 347 100	NO
rs58433430	NA	C/T	A_GTG	KLHL14	18	30 265 315	YES
rs2901215	NA	T/G	ACG_G	TIAL1	10	121 356 578	NO
rs56306258	NA	A/T	_CGTG	ANKRD12	18	9 136 756	NO
rs56033752	NA	G/A	AC_TG	DLG2	11	83 898 942	NO
rs34394782	NA	C/A/T	CACG_	CITED2	6	139 695 473	NO
rs34476700	NA	G/-	_CGTG	SENP3	17	7 463 409	NO
rs56298217	NA	G/A	GC_TG	CLK3	15	74 914 649	NO

The ‘\_’ symbol indicates the position of the SNP within the potential HIF-binding site. Max\_HBS indicates whether the SNP maps to the highest scoring HBS identified for each locus.

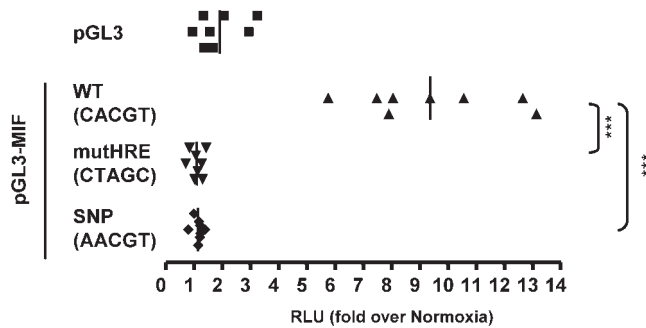
Note that only one of the possible alleles generates an RCGTG motif.

this point mutation had no effect on the basal activity of the promoter showing an activity  $1.16 \pm 0.46$  (mean  $\pm$  SD) over wild-type promoter ( $P = 0.85$ ). These results demonstrate that some polymorphic variants have a dramatic effect on gene regulation by hypoxia.

## DISCUSSION

The identification of the complete set of genes directly regulated by HIF is essential to fully understand the array of cellular responses activated during adaptation





**Figure 5.** The allelic variant C/A (rs17004038) abrogates MIF induction by hypoxia. HeLa cells transfected with a reporter plasmid containing MIF genomic region (−31 to +108) upstream a luciferase reporter gene. Where indicated the consensus HRE sequence (CACGT) was mutated to CTAGC (mutHRE) or to AACGT (SNP). The graph represents the corrected luciferase activity values of each construct in cells exposed to hypoxia over the luciferase activity obtained in normoxic cells. Data show the results for eight experiments and its mean value (vertical line). Statistically significant differences with control group (WT) are indicated by asterisks ( $***p < 0.001$ ).

to hypoxia. In this task, gene expression and TFBS data generated by high-throughput tools are fundamental. However, comparison between different studies (Figures 1A and 2A) reveals little overlap in the results, probably not only because of the particularities in the response to hypoxia in each particular system, but also because of the intrinsic noise associated to these techniques. Herein we describe a novel probabilistic strategy that integrates the rich information contained in gene expression profiling databases with classic bioinformatic approaches to predict TFBSs. The evaluation of this strategy, using published ChIP–chip data as a benchmark, indicates that it has a low error rate while retaining a high sensitivity. In agreement, our experimental validation revealed that five out of six of predicted targets were in fact true positives while 8/10 negatives were true negatives (Figures 3 and 4). Thus, the computational strategy described herein proved to be comparable in success rate to the experimental identification of HBSs by means of ChIP–chip, and it is hence an attractive alternative until these high-throughput techniques become more cost efficient.

In our strategy, the identification of relevant HBSs relies on the similitude of the potential HBS sequences to a PSSM that includes positions other than the core RCGTG. This PSSM was obtained by our analysis of a set of 46 sequences derived from well-characterized HREs (Supplementary Table S4). In contrast to our result, the analysis of genomic fragments bound by HIF failed to identify extended sequence preferences beyond the core RCGTG (28,29). Thus, it could be argued that the extended motif revealed by our analysis is consequence of the (relatively) reduced number of sequences in the alignment (46 sequences). However, using a PSSM based on this extended motif, we found that the HBSs identified by ChIP–chip had an associated score significantly higher than background sites (Figure 3B) and that, within a given locus, the functional HBS coincides with the highest scoring one (Figure 3C). These results strongly argue in favor of the PSSM-based score as a reliable parameter to discriminate functional HBSs and justify its inclusion in

our computational strategy. The information content of the extra conserved positions is low as compared to that of the core HRE (Supplementary Table S4), suggesting that probably each individual position plays a minor role on HIF sequence preference. However, its combined effect could explain the preferential binding of HIF to these sequences. Further work is required to prove the relevance of these conserved positions outside the core RCGTG and, in the event of them being relevant, to determine whether they are part of the HIF $\alpha$  (or HIF $\beta$ ) binding site, form the binding site of an unrelated transcription factor or just confer a favorable structure. Another premise in our strategy is that HREs are restricted to genomic sequences conserved during evolution. We imposed this restriction to our model knowing that not all *cis*-regulatory motifs are conserved (44). Nevertheless, this restriction was required to minimize the number of false positives while allowing a good sensitivity (~80%). In fact, this high sensitivity suggests that most real HBSs do in fact lie within conserved regions. In agreement, we found that 79% (254 out of 320) of the genomic fragments reported to bind HIF (29) contained one or more PhastCons elements. Thus, evolutionary conservation constraints are useful in reaching an optimum trade-off between sensitivity and specificity. In addition, our results imply that most (80%) of the experimentally identified HBSs are associated with CNSs. A further potential limitation of our strategy is imposed by the meta-analysis of gene expression profiling experiments. In our meta-analysis, genes showing a tissue-specific regulation by hypoxia, such as erythropoietin (EPO), fail to be identified as hypoxia regulated genes. In order to mitigate this effect, in our strategy, the contribution of the meta-analysis to the classification of a gene as an HIF target is weighted by the consistency of its regulation by hypoxia (*P*-value) across the panel of microarray experiments (see ‘Materials and Methods’ section for details). In fact, this correction led us to the identification of RCOR2 as a HIF target in spite of it being induced by hypoxia in a limited number of experiments. Our analysis correctly identified the HRE driving EPO expression, but this gene was not selected as a HIF target because its associated  $P_T/P_B$  value was 5.17, right below the threshold of 6.5. In fact, EPO ranked in position 360 out of 11 672 analyzed genes. The design of our strategy tries to minimize the false positives to give a highly reliable list of HIF-targets but, because of the restrictions imposed, several HIF targets are missed as is the case of EPO. Hence, the list of 217 HIF-targets reported herein is clearly an underestimation of the whole complement of genes regulated by HIF.

An unexpected conclusion from our results is the lack of HBS enrichment in hypoxia-downregulated genes (Table 1). In agreement with our statistical approach, a recent ChIP–chip study (28) also failed to find association between HIF binding and transcriptional downregulation. In fact, although a direct effect on gene downregulation has been documented for some genes, such as CAD (45), they are rare exceptions, being upregulation of targets the predominant effect upon HIF binding. Thus, it is tempting to speculate that, in contrast to gene induction, most of

the transcriptional downregulation triggered by hypoxia is either HIF independent or mediated by a secondary factor downstream of HIF (indirect effect). The lack of genes consistently downregulated by hypoxia in gene expression data sets (Figures 3 and 4) supports that HIF does not play a direct role in gene downregulation. Interestingly, among the HIF targets identified in our study, there are several factors involved in transcriptional repression, including the novel HIF target RCOR2 described herein (Figure 4). Thus, it is plausible that HIF indirectly promotes the transcriptional repression of specific genes by controlling the expression of co-repressors. However, other mechanisms could be envisioned to explain an indirect effect of HIF on gene downregulation. For example, it has been recently described that hypoxia/HIF leads to the induction of microRNAs (46) that, in turn, could lead to downregulation of specific groups of genes. Thus, much work is required to understand the molecular mechanisms responsible for hypoxia-induced gene repression.

An important feature of our strategy is that it is not restricted to a particular HIF isoform. Most of the GEO data sets used for the meta-analysis (Supplementary Table S2) derive from experiments that used hypoxia or the EGLN inhibitor DMOG as stimuli and thus activated all HIF $\alpha$  subunits present in the cells. Only in two tables (GSE2020), a specific isoform was activated by overexpression. On the other hand, it is assumed that HIF1 $\alpha$  and HIF2 $\alpha$  bind to the same motif (RCGTG) and that their differential target preference stems from isoform-specific cooperation with other transcription factors (47,48). In fact, the binding of both isoforms to a common motif was recently confirmed by comparison of the genomic sequences immunoprecipitated with HIF1 $\alpha$  and HIF2 $\alpha$  (28). Thus, the approaches used in our strategy are not biased toward the preferential identification of isoform-specific targets. In fact, the list of candidates (Supplementary Table S6) includes genes reported as HIF1 $\alpha$  [BNIP3, (48)] and HIF2 $\alpha$  specific [CITED2, (47)].

The precise identification of HBSs did not only lead to the identification of direct HIF targets but also allowed us to predict polymorphisms that could affect gene regulation by hypoxia. In this work, we identified several SNPs mapping to predicted HBSs and demonstrated, in the case of MIF promoter, that specific allelic variants result in a severely impaired response to hypoxia. Thus, individuals presenting this variant probably fail to properly upregulate MIF in response to hypoxia. To our knowledge, these results constitute the first demonstration that the response to hypoxia could vary slightly between individuals of a population. It is difficult to predict the physiological consequences of the lack of hypoxic induction of MIF and further work is necessary to address this question. However, it is likely that the abrogation of hypoxic gene induction had dramatic consequences. In agreement, elimination of the HRE driving the hypoxic upregulation of VEGFA leads to motor neuron degeneration (49) and abnormal retinal neovascularization (50). Given the number of pathologies that course with hypoxia, our results point to a potential source of variability in the clinical course and/or response to treatments

among different individuals. In addition, these results support the hypothesis that mutations in regulatory regions, rather than in coding sequences, are important to explain inter-individual variation. With the completion of ongoing sequencing projects aimed at the identification of novel SNPs, we foresee that the number of variants affecting HBSs will increase.

In conclusion, the data presented herein demonstrate that integration of gene expression profiling and *in silico* identification of TFBSs is a successful approach for the identification of direct target genes. In agreement, during the writing of our manuscript, a report was published (51) that employs a similar strategy to identify HIF targets. Interestingly, although both works identify a list of about 200 HIF targets, there is little overlap in the identity of the individual target genes (only 37 genes were coincident, see Supplementary Table S8), reflecting important differences in the approaches followed in each work. The application of our strategy led to the identification of a set of novel (potential) HIF targets and our experimental validation demonstrated the reliability of these predictions. Moreover, we have found that, at least one of the predicted targets, RCOR2, is an HIF target gene regulated by hypoxia. In addition, elsewhere we demonstrate that two additional novel targets, GYS1 and UGP2, are also regulated by hypoxia in an HIF-dependent manner (manuscript submitted for publication). Finally, we identified polymorphisms mapping to our predicted HBSs and demonstrated that specific alleles have a profound impact on the regulation of transcription by hypoxia. Altogether, these results expand our understanding of the adaptive responses to hypoxia and suggest, for the first time, that this response can vary among individuals.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank Gema Moreno (Department of Biochemistry, Universidad Autónoma de Madrid, Madrid, Spain) for critical reading of the manuscript and Benjamin A.T. Rodríguez (Human Cancer Genetics Comprehensive Cancer Center, The Ohio State University, Columbus, OH, USA) for his valuable suggestions regarding validation of ChIP by qPCR. We also thank Dr. Manuel O. Landazuri and the researchers in his team for their kind support and help in many aspects of this work.

## FUNDING

Ministerio de Ciencia y Tecnología/Ministerio de Ciencia e Innovación (SAF2005-00180 and SAF2008-03147), Comunidad Autónoma de Madrid (S-SAL-0311\_2006); and the METOXIA, project ref. HEALTH-F2-2009-222741, under the 7th Research Framework Programme of the European Union. Funding for open access charge: SAF2008-03147.

*Conflict of interest statement.* None declared.

## REFERENCES

- Wenger, R.H., Stiehl, D.P. and Camenisch, G. (2005) Integration of oxygen signaling at the consensus HRE. *Sci. STKE*, **2005**, re12.
- Semenza, G.L. and Wang, G.L. (1992) A nuclear factor induced by hypoxia via de novo protein synthesis binds to the human erythropoietin gene enhancer at a site required for transcriptional activation. *Mol. Cell. Biol.*, **12**, 5447–5454.
- Salceda, S., Beck, I. and Caro, J. (1996) Absolute requirement of aryl hydrocarbon receptor nuclear translocator protein for gene activation by hypoxia. *Arch. Biochem. Biophys.*, **334**, 389–394.
- Salceda, S. and Caro, J. (1997) Hypoxia-inducible factor 1 $\alpha$  (HIF-1 $\alpha$ ) protein is rapidly degraded by the ubiquitin-proteasome system under normoxic conditions. Its stabilization by hypoxia depends on redox-induced changes. *J. Biol. Chem.*, **272**, 22642–22647.
- Jiang, B.H., Zheng, J.Z., Leung, S.W., Roe, R. and Semenza, G.L. (1997) Transactivation and inhibitory domains of hypoxia-inducible factor 1 $\alpha$ . Modulation of transcriptional activity by oxygen tension. *J. Biol. Chem.*, **272**, 19253–19260.
- Ivan, M., Kondo, K., Yang, H., Kim, W., Valiando, J., Ohh, M., Salic, A., Asara, J.M., Lane, W.S. and Kaelin, W.G. Jr (2001) HIF1 $\alpha$  targeted for VHL-mediated destruction by proline hydroxylation: implications for O<sub>2</sub> sensing. *Science*, **292**, 464–468.
- Jaakkola, P., Mole, D.R., Tian, Y.M., Wilson, M.I., Gielbert, J., Gaskell, S.J., Kriegsheim, A., Hebestreit, H.F., Mukherji, M., Schofield, C.J. *et al.* (2001) Targeting of HIF-1 $\alpha$  to the von Hippel–Lindau ubiquitylation complex by O<sub>2</sub>-regulated prolyl hydroxylation. *Science*, **292**, 468–472.
- Bruick, R.K. and McKnight, S.L. (2001) A conserved family of prolyl-4-hydroxylases that modify HIF. *Science*, **294**, 1337–1340.
- Epstein, A.C., Gleadle, J.M., McNeill, L.A., Hewitson, K.S., O'Rourke, J., Mole, D.R., Mukherji, M., Metzen, E., Wilson, M.I., Dhanda, A. *et al.* (2001) C. elegans EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. *Cell*, **107**, 43–54.
- Maxwell, P.H., Wiesener, M.S., Chang, G.W., Clifford, S.C., Vaux, E.C., Cockman, M.E., Wykoff, C.C., Pugh, C.W., Maher, E.R. and Ratcliffe, P.J. (1999) The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature*, **399**, 271–275.
- Hewitson, K.S., McNeill, L.A., Riordan, M.V., Tian, Y.M., Bullock, A.N., Welford, R.W., Elkins, J.M., Oldham, N.J., Shoumou, B., Gleadle, J.M. *et al.* (2002) Hypoxia-inducible factor (HIF) asparagine hydroxylase is identical to factor inhibiting HIF (FIH) and is related to the cupin structural family. *J. Biol. Chem.*, **277**, 26351–26355.
- Lando, D., Peet, D.J., Gorman, J.J., Whelan, D.A., White, M.F. and Bruick, R.K. (2002) FIH-1 is an asparaginyl hydroxylase enzyme that regulates the transcriptional activity of hypoxia-inducible factor. *Genes Dev.*, **16**, 1466–1471.
- Lando, D., Peet, D.J., Whelan, D.A., Gorman, J.J. and Whitelaw, M.L. (2002) Asparagine hydroxylation of the HIF transactivation domain a hypoxic switch. *Science*, **295**, 858–861.
- Hu, C.-J., Wang, L.-Y., Chodosh, L.A., Keith, B. and Simon, M.C. (2003) Differential roles of hypoxia-inducible factor 1 $\alpha$  (HIF-1 $\alpha$ ) and HIF-2 $\alpha$  in hypoxic gene regulation. *Mol. Cell. Biol.*, **23**, 9361–9374.
- Raval, R.R., Lau, K.W., Tran, M.G.B., Sowter, H.M., Mandriota, S.J., Li, J.-L., Pugh, C.W., Maxwell, P.H., Harris, A.L. and Ratcliffe, P.J. (2005) Contrasting properties of hypoxia-inducible factor 1 (HIF-1) and HIF-2 in von Hippel–Lindau-associated renal cell carcinoma. *Mol. Cell. Biol.*, **25**, 5675–5686.
- Papandreou, I., Cairns, R.A., Fontana, L., Lim, A.L. and Denko, N.C. (2006) HIF-1 mediates adaptation to hypoxia by actively downregulating mitochondrial oxygen consumption. *Cell Metab.*, **3**, 187–197.
- Kim, J.-W., Tchernyshyov, I., Semenza, G.L. and Dang, C.V. (2006) HIF-1-mediated expression of pyruvate dehydrogenase kinase: a metabolic switch required for cellular adaptation to hypoxia. *Cell Metab.*, **3**, 177–185.
- Bostrom, P., Magnusson, B., Svensson, P.-A., Wiklund, O., Boren, J., Carlsson, L.M.S., Stahlman, M., Olofsson, S.-O. and Hulten, L.M. (2006) Hypoxia converts human macrophages into triglyceride-loaded foam cells. *Arterioscler. Thromb. Vasc. Biol.*, **26**, 1871–1876.
- Elvidge, G.P., Glenny, L., Appelhoff, R.J., Ratcliffe, P.J., Ragoussis, J. and Gleadle, J.M. (2006) Concordant regulation of gene expression by hypoxia and 2-oxoglutarate-dependent dioxygenase inhibition. *J. Biol. Chem.*, **281**, 15215–15226.
- Kasper, L.H., Boussouar, F., Boyd, K., Xu, W., Biesen, M., Rehg, J., Baudino, T.A., Cleveland, J.L. and Brindle, P.K. (2005) Two transactivation mechanisms cooperate for the bulk of HIF-1-responsive gene expression. *EMBO J.*, **24**, 3846–3858.
- Allen, J.W., Khetani, S., Johnson, R. and Bhatia, S. (2006) In vitro liver tissue model established from transgenic mice: role of HIF-1 $\alpha$  on hypoxic gene expression. *Tissue Engg.*, **12**, 3135–3147.
- Mense, S.M., Sengupta, A., Zhou, M., Lan, C., Bentsman, G., Volsky, D.J. and Zhang, L. (2006) Gene expression profiling reveals the profound upregulation of hypoxia-responsive genes in primary human astrocytes. *Physiol. Genomics*, **25**, 435–449.
- Ray, J.B., Arab, S., Deng, Y., Liu, P., Penn, L., Courtman, D.W. and Ward, M.E. (2008) Oxygen regulation of arterial smooth muscle cell proliferation and survival. *Am. J. Physiol. Heart Circ. Physiol.*, **294**, H839–H852.
- Wang, V., Davis, D.A., Haque, M., Huang, L.E. and Yarchoan, R. (2005) Differential gene up-regulation by hypoxia-inducible factor-1 $\alpha$  and hypoxia-inducible factor-2 $\alpha$  in HEK293T cells. *Cancer Res.*, **65**, 3299–3306.
- Guimbellot, J., Erickson, S., Mehta, T., Wen, H., Page, G., Sorscher, E. and Hong, J. (2009) Correlation of microRNA levels during hypoxia with predicted target mRNAs through genome-wide microarray analysis. *BMC Med. Genomics*, **2**, 15.
- Irigoyen, M., Ansó, E., Martínez, E., Garayoa, M., Martínez-Irujo, J.J. and Rouzaut, A. (2007) Hypoxia alters the adhesive properties of lymphatic endothelial cells. A transcriptional and functional study. *Biochim. Biophys. Acta – Mol. Cell. Res.*, **1773**, 880–890.
- Cahan, P., Rovegno, F., Mooney, D., Newman, J.C., St. Laurent, G. III and McCaffrey, T.A. (2007) Meta-analysis of microarray results: challenges, opportunities, and recommendations for standardization. *Gene*, **401**, 12–18.
- Mole, D.R., Blancher, C., Copley, R.R., Pollard, P.J., Gleadle, J.M., Ragoussis, J. and Ratcliffe, P.J. (2009) Genome-wide association of hypoxia-inducible factor (HIF)-1 $\alpha$  and HIF-2 $\alpha$  DNA binding with expression profiling of hypoxia-inducible transcripts. *J. Biol. Chem.*, **284**, 16767–16775.
- Xia, X., Lemieux, M.E., Li, W., Carroll, J.S., Brown, M., Liu, X.S. and Kung, A.L. (2009) Integrative analysis of HIF binding and transactivation reveals its role in maintaining histone methylation homeostasis. *Proc. Natl Acad. Sci. USA*, **106**, 4260–4265.
- Elnitski, L., Jin, V.X., Farnham, P.J. and Jones, S.J.M. (2006) Locating mammalian transcription factor binding sites: A survey of computational and experimental techniques. *Genome Res.*, **16**, 1455–1464.
- Pescador, N., Cuevas, Y., Naranjo, S., Alcaide, M., Villar, D., Landázuri, M.O. and del Peso, L. (2005) Regulation of the egl nine homologue 3 (egl3/phd3) gene: Identification of a functional hypoxia-responsive element. *Biochem. J.*, **390**, 189–197.
- Kimura, H., Weisz, A., Ogura, T., Hitomi, Y., Kurashima, Y., Hashimoto, K., D'Acquisto, F., Makuuchi, M. and Esumi, H. (2001) Identification of hypoxia-inducible factor-1 (HIF-1) ancillary sequence and its function in vascular endothelial growth factor gene induction by hypoxia and nitric oxide. *J. Biol. Chem.*, **276**, 2292–2298.
- Gilligan, P., Brenner, S. and Venkatesh, B. (2002) Fugu and human sequence comparison identifies novel human genes and conserved non-coding sequences. *Gene*, **294**, 35–44.
- Nobrega, M.A. and Pennacchio, L.A. (2004) Comparative genomic analysis as a tool for biological discovery. *J. Physiol.*, **554**, 31–39.

35. Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S. *et al.* (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.*, **15**, 1034–1050.
36. Bussemaker, H.J., Li, H. and Siggia, E.D. (2001) Regulatory element detection using correlation with expression. *Nat. Genet.*, **27**, 167–174.
37. Hoek, K.S., Schlegel, N.C., Eichhoff, O.M., Widmer, D.S., Praetorius, C., Einarsson, S.O., Valgeirsdottir, S., Bergsteinsdottir, K., Schepsky, A., Dummer, R. *et al.* (2008) Novel MITF targets identified using a two-step DNA microarray strategy. *Pigment Cell Melanoma Res.*, **21**, 665–676.
38. Jeffery, I.B., Madden, S.F., McGettigan, P.A., Perriere, G., Culhane, A.C. and Higgins, D.G. (2007) Integrating transcription factor binding site information with gene expression datasets. *Bioinformatics*, **23**, 298–305.
39. Edgar, R., Domrachev, M. and Lash, A.E. (2002) Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.*, **30**, 207–210.
40. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. and Haussler, A.D. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.
41. Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D. and Kent, W.J. (2004) The UCSC table browser data retrieval tool. *Nucleic Acids Res.*, **32**, D493–D496.
42. Wood, S.M., Gleadle, J.M., Pugh, C.W., Hankinson, O. and Ratcliffe, P.J. (1996) The role of the aryl hydrocarbon receptor nuclear translocator (ARNT) in hypoxic induction of gene expression. Studies in ARNT-deficient cells. *J. Biol. Chem.*, **271**, 15117–15123.
43. Baugh, J.A., Gantier, M., Li, L., Byrne, A., Buckley, A. and Donnelly, S.C. (2006) Dual regulation of macrophage migration inhibitory factor (MIF) expression in hypoxia by CREB and HIF-1. *Biochem. Biophys. Res. Comm.*, **347**, 895–903.
44. Alonso, E., Pernaute, B., Crespo, M., Gómez-Skarmeta, J.L. and Manzanares, M. (2008) Understanding the regulatory genome. *Int. J. Dev. Biol.*, **53**, 1367–1378.
45. Chen, K.-F., Lai, Y.-Y., Sun, H.S. and Tsai, S.-J. (2005) Transcriptional repression of human cad gene by hypoxia inducible factor-1 $\alpha$ . *Nucleic Acids Res.*, **33**, 5190–5198.
46. Kulshreshtha, R., Ferracin, M., Wojcik, S., Garzon, R., Alder, H., Agosto-Perez, F., Davuluri, R., Liu, C., Croce, C. and Negrini, M. (2007) A microRNA signature of hypoxia. *Mol. Cell. Biol.*, **27**, 1859–1867.
47. Hu, C.-J., Iyer, S., Sataur, A., Covelto, K.L., Chodosh, L.A. and Simon, M.C. (2006) Differential regulation of the transcriptional activities of hypoxia-inducible factor 1 alpha (HIF-1 $\alpha$ ) and HIF-2 $\alpha$  in stem cells. *Mol. Cell. Biol.*, **26**, 3514–3526.
48. Aprelikova, O., Wood, M., Tackett, S., Chandramouli, G.V.R. and Barrett, J.C. (2006) Role of ETS transcription factors in the hypoxia-inducible factor-2 target gene selection. *Cancer Res.*, **66**, 5641–5647.
49. Oosthuysen, B., Moons, L., Storkebaum, E., Beck, H., Nuyens, D., Brusselmans, K., Dorpe, J.V., Hellings, P., Gorselink, M., Heymans, S. *et al.* (2001) Deletion of the hypoxia-response element in the vascular endothelial growth factor promoter causes motor neuron degeneration. *Nat. Genet.*, **28**, 131–138.
50. Viores, S.A., Xiao, W.-H., Aslam, S., Shen, J., Oshima, Y., Nambu, H., Liu, H., Carmeliet, P. and Campochiaro, P.A. (2006) Implication of the hypoxia response element of the VEGF promoter in mouse models of retinal and choroidal neovascularization, but not retinal vascular development. *J. Cell. Physiol.*, **206**, 749–758.
51. Benita, Y., Kikuchi, H., Smith, A.D., Zhang, M.Q., Chung, D.C. and Xavier, R.J. (2009) An integrative genomics approach identifies hypoxia inducible factor-1 (HIF-1)-target genes that form the core response to hypoxia. *Nucleic Acids Res.*, **37**, 4587–4602.

# Hypoxia Promotes Glycogen Accumulation through Hypoxia Inducible Factor (HIF)-Mediated Induction of Glycogen Synthase 1

Nuria Pescador<sup>1,2</sup>, Diego Villar<sup>1,2,3</sup>, Daniel Cifuentes<sup>3,9a</sup>, Mar Garcia-Rocha<sup>3</sup>, Amaya Ortiz-Barahona<sup>1,2</sup>, Silvia Vazquez<sup>4</sup>, Angel Ordoñez<sup>6</sup>, Yolanda Cuevas<sup>1,2</sup>, David Saez-Morales<sup>5</sup>, Maria Laura Garcia-Bermejo<sup>5</sup>, Manuel O. Landazuri<sup>4</sup>, Joan Guinovart<sup>3</sup>, Luis del Peso<sup>1,2\*</sup>

**1** Departamento de Bioquímica, Universidad Autónoma de Madrid, Madrid, Spain, **2** Instituto de Investigaciones Biomédicas "Alberto Sols", Consejo Superior de Investigaciones Científicas, Madrid, Spain, **3** Institute for Research in Biomedicine (IRB Barcelona), University of Barcelona, Barcelona, Spain, **4** Department of Biochemistry and Molecular Biology, University of Barcelona, Barcelona, Spain, **5** Servicio de Inmunología Hospital de la Princesa, Universidad Autónoma de Madrid, Madrid, Spain, **6** Servicio de Anatomía Patológica Hospital Ramón y Cajal, Madrid, Spain

## Abstract

When oxygen becomes limiting, cells reduce mitochondrial respiration and increase ATP production through anaerobic fermentation of glucose. The Hypoxia Inducible Factors (HIFs) play a key role in this metabolic shift by regulating the transcription of key enzymes of glucose metabolism. Here we show that oxygen regulates the expression of the muscle glycogen synthase (GYS1). Hypoxic GYS1 induction requires HIF activity and a Hypoxia Response Element within its promoter. GYS1 gene induction correlated with a significant increase in glycogen synthase activity and glycogen accumulation in cells exposed to hypoxia. Significantly, knockdown of either HIF1 $\alpha$  or GYS1 attenuated hypoxia-induced glycogen accumulation, while GYS1 overexpression was sufficient to mimic this effect. Altogether, these results indicate that GYS1 regulation by HIF plays a central role in the hypoxic accumulation of glycogen. Importantly, we found that hypoxia also upregulates the expression of UTP:glucose-1-phosphate uridylyltransferase (UGP2) and 1,4- $\alpha$  glucan branching enzyme (GBE1), two enzymes involved in the biosynthesis of glycogen. Therefore, hypoxia regulates almost all the enzymes involved in glycogen metabolism in a coordinated fashion, leading to its accumulation. Finally, we demonstrated that abrogation of glycogen synthesis, by knock-down of GYS1 expression, impairs hypoxic preconditioning, suggesting a physiological role for the glycogen accumulated during chronic hypoxia. In summary, our results uncover a novel effect of hypoxia on glucose metabolism, further supporting the central importance of metabolic reprogramming in the cellular adaptation to hypoxia.

**Citation:** Pescador N, Villar D, Cifuentes D, Garcia-Rocha M, Ortiz-Barahona A, et al. (2010) Hypoxia Promotes Glycogen Accumulation through Hypoxia Inducible Factor (HIF)-Mediated Induction of Glycogen Synthase 1. PLoS ONE 5(3): e9644. doi:10.1371/journal.pone.0009644

**Editor:** Laszlo Tora, Institute of Genetics and Molecular and Cellular Biology, France

**Received:** October 13, 2009; **Accepted:** January 27, 2010; **Published:** March 12, 2010

**Copyright:** © 2010 Pescador et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by grants from Ministerio de Ciencia y Tecnología/Ministerio de Ciencia e Innovación (SAF2005-00180 and SAF2008-03147 to L. P. and SAF2004-0824 to M. O. L.), Red Cardiovascular (RECAVA) and Comunidad Autónoma de Madrid (S-SAL-0311\_2006). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: luis.peso@uam.es

9 These authors contributed equally to this work.

a Current address: Yale Medical School, Yale University, New Haven, Connecticut, United States of America

## Introduction

Aerobic glycolysis yields sixteen times more ATP than glucose fermentation. Thus, metazoans are heavily dependent on the highly energy-efficient aerobic metabolism to meet their large ATP demands. Since oxygen is the final electron acceptor during mitochondrial respiration, this element is essential for metazoan metabolism. In fact, most of the oxygen consumed by animals is required to sustain oxidative phosphorylation. However, several animal species are able to adapt to reduced oxygen tensions during variable periods of time [1]. These include diving animals, fossorial mammals, animals living at high altitude and species of fish that live in stagnant waters, among others [1]. In nearly all these cases, the adaptation to hypoxia involves systemic and cellular responses aimed to reduce oxygen consumption and optimize its utilization.

At the cellular level, the reduction in oxygen consumption is mediated by a shift from oxidative to fermentative glucose metabolism [2,3,4,5], the fine-tuning of mitochondrial respiration [6] and an adjustment of mitochondria number. This hypometabolic state is achieved by the induction of a specific gene expression program under the control of the Hypoxia-Inducible Factor (HIF) family of transcription factors. This family comprises three transcription factors (HIF-1, HIF-2 and HIF-3) that are heterodimers of a constitutively expressed  $\beta$  subunit and an oxygen-sensitive  $\alpha$  subunit (HIF1 $\alpha$ , HIF2 $\alpha$  or HIF3 $\alpha$ ) [7,8,9]. In the presence of oxygen, HIF $\alpha$  is extremely unstable due to an oxygen-dependent posttranslational hydroxylation that targets it for proteosomal degradation [10,11,12,13,14]. A family of prolyl hydroxylases (EGLNs), that require molecular oxygen as a reaction cosubstrate, catalyzes the hydroxylation of HIF $\alpha$  [15,16,17]. Hence, under hypoxia, hydroxylation is compromised leading to

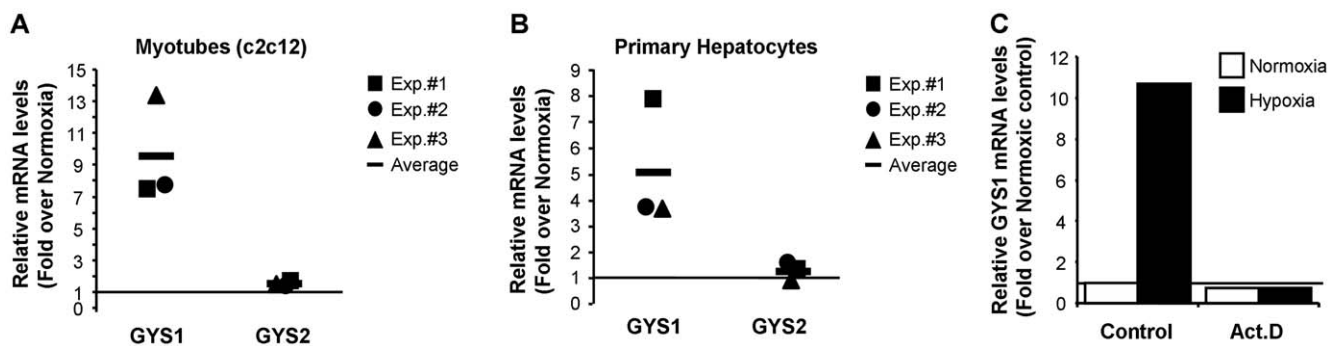
HIF stabilization and, subsequently, to its binding the hypoxia-response element (RCGTG motif) on the regulatory region of target genes [18]. Among these genes are those of glucose transporter SLC2A1 (Glut1) and almost all glycolytic enzymes [2], whose induction leads to an increased glycolytic flux. In addition, hypoxia regulates the end product of glycolysis, pyruvate, which is actively diverted from mitochondrial oxidation due to the HIF-mediated induction of Pyruvate Dehydrogenase Kinases (PDK1 and PDK4) [3,4,19]. Finally, HIF also induces the transcription of lactate dehydrogenase (LDH) [20] and lactate transporter (MCT4) [21], whose activities account for the transformation of pyruvate into lactate and its export out of the cell, respectively. Overall, the coordinated induction of all these genes by HIF decreases glucose oxidation, contributing to reduced oxygen consumption, and increases anaerobic ATP production to maintain energy balance. Thus, the adaptation to hypoxia is critically dependent on the reprogramming of glucose metabolism and, accordingly, a large number of HIF-target genes function in coordination to promote this metabolic shift. Importantly, because of all these metabolic adaptations, hypoxic cells rely heavily on an appropriate glucose supply. To ensure glucose availability, cells that depend on anaerobic glycolysis to sustain ATP production, such as skeletal myotubes, accumulate large amounts of glucose in the form of glycogen. In agreement, some reports described an increase in glycogen upon exposure of cells to hypoxia [22,23] or hypoxia-mimetics [24]. However, in spite of these pieces of evidence, the effect of hypoxia on glycogen metabolism and the molecular mechanism involved are yet to be addressed.

Herein we demonstrate that hypoxia leads to glycogen accumulation through the coordinated regulation of glycogen biosynthesis. We show that GYS1 is a novel hypoxia-inducible, HIF-dependent gene, whose induction mediates glycogen accumulation following oxygen deprivation. Consequently, this work reveals a novel metabolic response to low oxygen and emphasizes the relevance of glucose metabolism in the adaptation to hypoxia. Furthermore, we found that GYS1 interference abrogates hypoxic preconditioning in an *in vitro* model of ischemia. We hence propose that the HIF-dependent increase in glycogen stores is a hypoxia-triggered adaptation that prepares cells to cope with further oxygen restrictions by ensuring adequate substrate supply for anaerobic glycolysis.

## Results

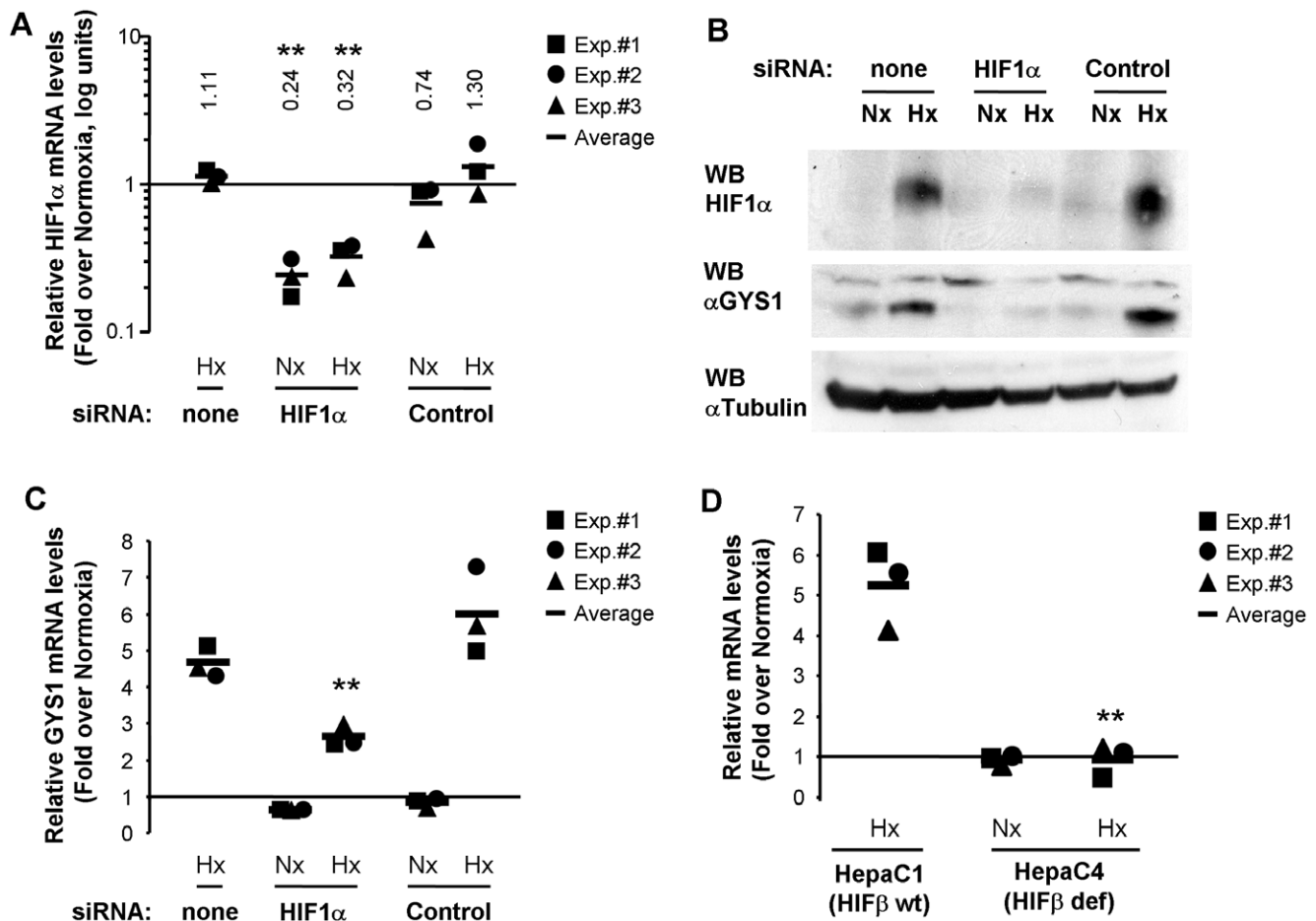
### Hypoxia induces GYS1 expression in a HIF-dependent manner

With the aim of understanding the whole array of metabolic adaptations that cells employ to cope with reduced oxygen tension, we searched the human genome for the occurrence of potential hypoxia-response elements (HREs). This search was based on the evolutionary conservation of HREs and their similitude to known functional HRE-containing regulatory regions [25]. The details of the strategy were reported elsewhere (Ortiz et al. *Nucl. Acids Res. in press*). One of the potential HREs identified by this approach was located in the promoter of the muscle isoform of glycogen synthase, GYS1. To investigate the potential regulation of GYS1 by hypoxia, we exposed c2c12 skeletal muscle myotubes to 1% oxygen and determined the level of GYS1 mRNA by quantitative PCR. Figure 1A shows that GYS1 mRNA was strongly induced by hypoxia in c2c12 cells. In contrast, hypoxia did not alter the mRNA levels of the closely related GYS2 (liver glycogen synthase) enzyme (figure 1A). In agreement, we did not find evolutionarily conserved HREs in the GYS2 locus. The hypoxic induction of GYS1 was not restricted to differentiated skeletal muscle cells, but was also observed in myoblast as well as other primary and stable cell lines (Figure S1 and data not shown). It is noteworthy that hypoxia also led to GYS1 induction in the hepatoma cell lines HepG2 (data not shown) and HepaC1 (see below), as well as in primary hepatocytes (figure 1B). The induction of GYS1 mRNA was inhibited by actinomycin D treatment (figure 1C), suggesting that hypoxia affected its transcription rather than its stability. Altogether, these results indicated that GYS1 was a hypoxia-inducible gene, and strongly hinted that it was also a HIF-target gene. In agreement, treatment of c2c12 cells with a siRNA targeting HIF1 $\alpha$  (figure 2A and 2B) reduced GYS1 induction by hypoxia (figure 2B and C). The partial induction of GYS1 by hypoxia observed in HIF1 $\alpha$  knock down cells could be explained by the remaining HIF1 $\alpha$  expression and/or the contribution of HIF2 $\alpha$  to GYS1 regulation. To corroborate the implication of HIF in this process, we investigated the effect of hypoxia on GYS1 expression in the HepaC4 HIF $\beta$ -deficient cells (figure 2D). These cells contain a mutant HIF $\beta$  subunit with a severely reduced half life and, as a consequence, they are deficient in HIF activity



**Figure 1. GYS1, but not GYS2, is a hypoxia-regulated HIF target gene.** c2c12 myotubes (A) or primary mouse hepatocytes (B) were exposed to normoxia or hypoxia for 12 hours and the level of GYS1 and GYS2 mRNA was determined by quantitative PCR. The amount of each mRNA in samples was normalized to the content of  $\beta$ -actin mRNA in the same sample. The graph represents the fold values of hypoxic over normoxic mRNA levels normalized to the value of 1 (horizontal line). Data shown are the results of three independent experiments and their mean. (C) c2c12 cells were treated with 1  $\mu$ g/ml actinomycin D (Act.D) or vehicle (control) for 30 minutes and then exposed to normoxia or 1% oxygen (Hypoxia) for 12 hours. The graph represents GYS1 mRNA levels in each condition relative to the untreated (control) normoxic cells. The experiment was repeated twice with identical results.

doi:10.1371/journal.pone.0009644.g001



**Figure 2. HIF mediates GYS1 induction by hypoxia.** c2c12 myoblasts were transfected with siRNA against mouse HIF1 $\alpha$  (HIF1 $\alpha$ ), an siRNA containing an irrelevant sequence (Scramble) or left untreated (none). 24 hours after transfection cells were cultured for 12 hours at 21% (Nx) or 1% oxygen (Hx) and then processed to determine mRNA (A,C), or protein (B) levels. Graphs represent normalized mRNA levels in each condition relative to the untreated normoxic cells. Data shown are the results of three independent experiments and their mean. \*\*,  $p < 0.01$  as compared to samples transfected with irrelevant siRNA (Control) and exposed to the same oxygen tension. Numbers represent average values. (D) HIF-competent (HIF $\beta$  wt), Hepa C1, and HIF-deficient (HIF $\beta$  def), Hepa C4, cells were exposed to normoxia (Nx) or 1% oxygen (Hx) for 12 hours and processed to determine GYS1 mRNA level. The graph represents the normalized GYS1 mRNA levels in each condition relative to the normoxic Hepa C1 cells (horizontal line) from three independent experiments. \*\*,  $p < 0.01$  as compared to HIF-competent cells exposed to the same oxygen tension. doi:10.1371/journal.pone.0009644.g002

[26,27]. As shown in figure 2D, while hypoxia induced glycogen synthase accumulation in the parental HIF-competent HepaC1 cells, this effect was completely blunted in the HepaC4 cells (figure 2D). Taken together, these results identify GYS1 as a novel HIF-target gene upregulated by hypoxia.

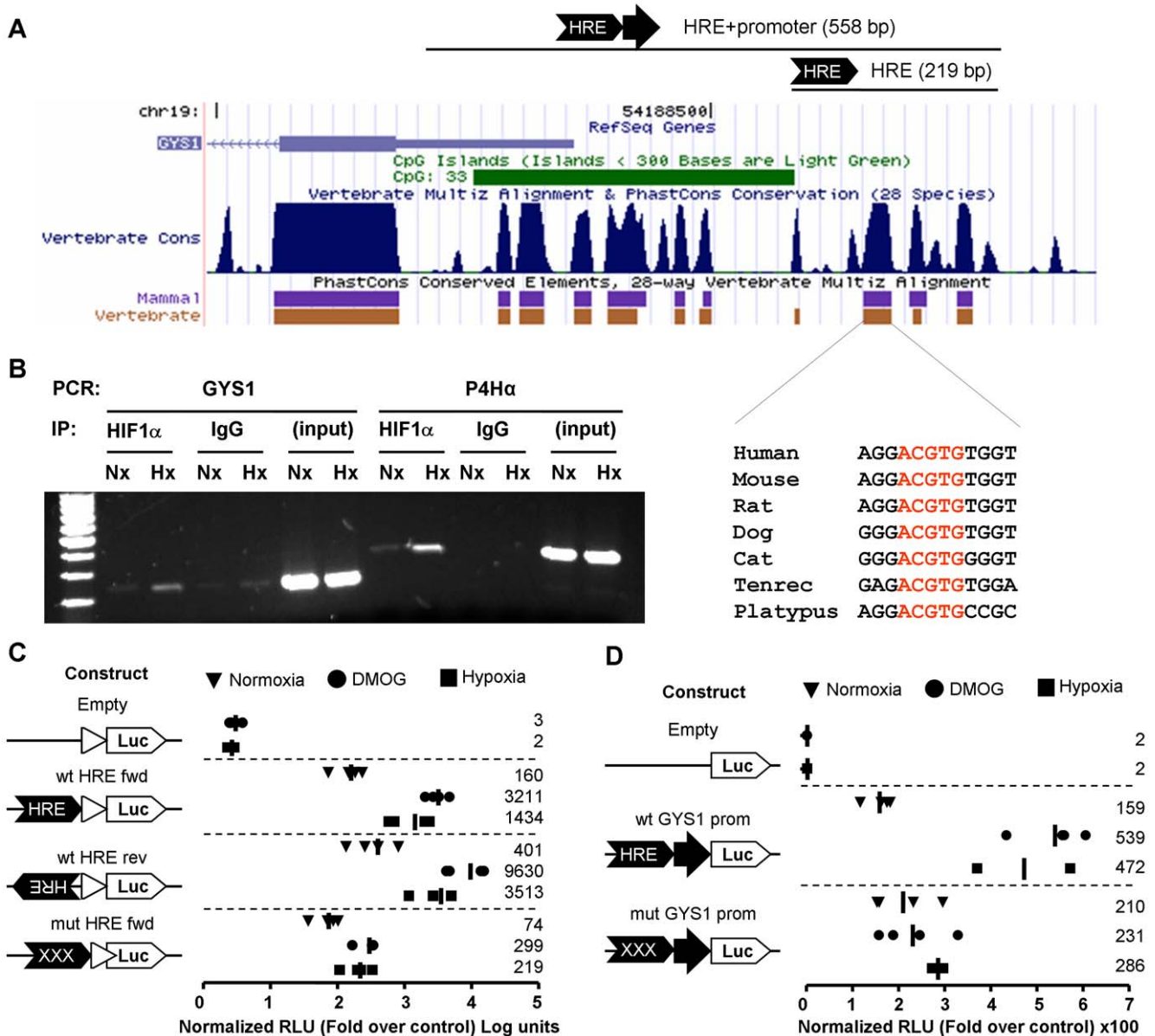
### GYS1 promoter region contains a functional HRE

The HRE identified by our bioinformatic search (Ortiz et al. Nucl. Acids Res. *in press*) was located at position -314 in the promoter region of human GYS1 (figure 3A). This site matched the canonical ACGTG motif and was conserved across mammals (figure 3A). Moreover, the HRE motif was included in a non-coding region conserved across vertebrates (figure 3A, PhastCons vertebrate track) [28]. Importantly, HIF1 $\alpha$  associated with the genomic region containing this HRE motif upon exposure of c2c12 cells to hypoxia (Figure 3B), suggesting that, in intact cells, HIF binds this site. To test the functionality of this HRE, we cloned the region from -211 to -429 upstream a minimal promoter driving the expression of the firefly luciferase reporter gene. As shown in figure 3C, hypoxia and the EGLN inhibitor DMOG strongly induced luciferase activity in cells transfected with the

construct containing the potential HRE regardless of its orientation. Importantly, mutation of the HRE motif completely impaired the induction of this construct by hypoxia (figure 3C). Next, we tested whether this element was able to confer hypoxia-inducibility to the native GYS1 promoter. To this end, we cloned the whole promoter region of the human GYS1 gene including the HRE (+128 to -429) into the pGL3-basic plasmid. This genomic region showed promoter activity since it drove luciferase expression above pGL3 basal levels (figure 3D). Significantly, the promoter activity of this construct was robustly induced by hypoxia and DMOG (figure 3D). In addition, the induction by Hypoxia and DMOG, but not the basal promoter activity, was lost when the reporter construct included a mutation in the HRE motif (figure 3D). In summary, GYS1 contains an evolutionarily conserved HRE located at -314 that mediates its induction by hypoxia.

### Hypoxia-mediated induction of GYS1 results in increased glycogen synthase activity

The experiments described above demonstrated that hypoxia induced GYS1 transcription in a HIF-dependent manner.

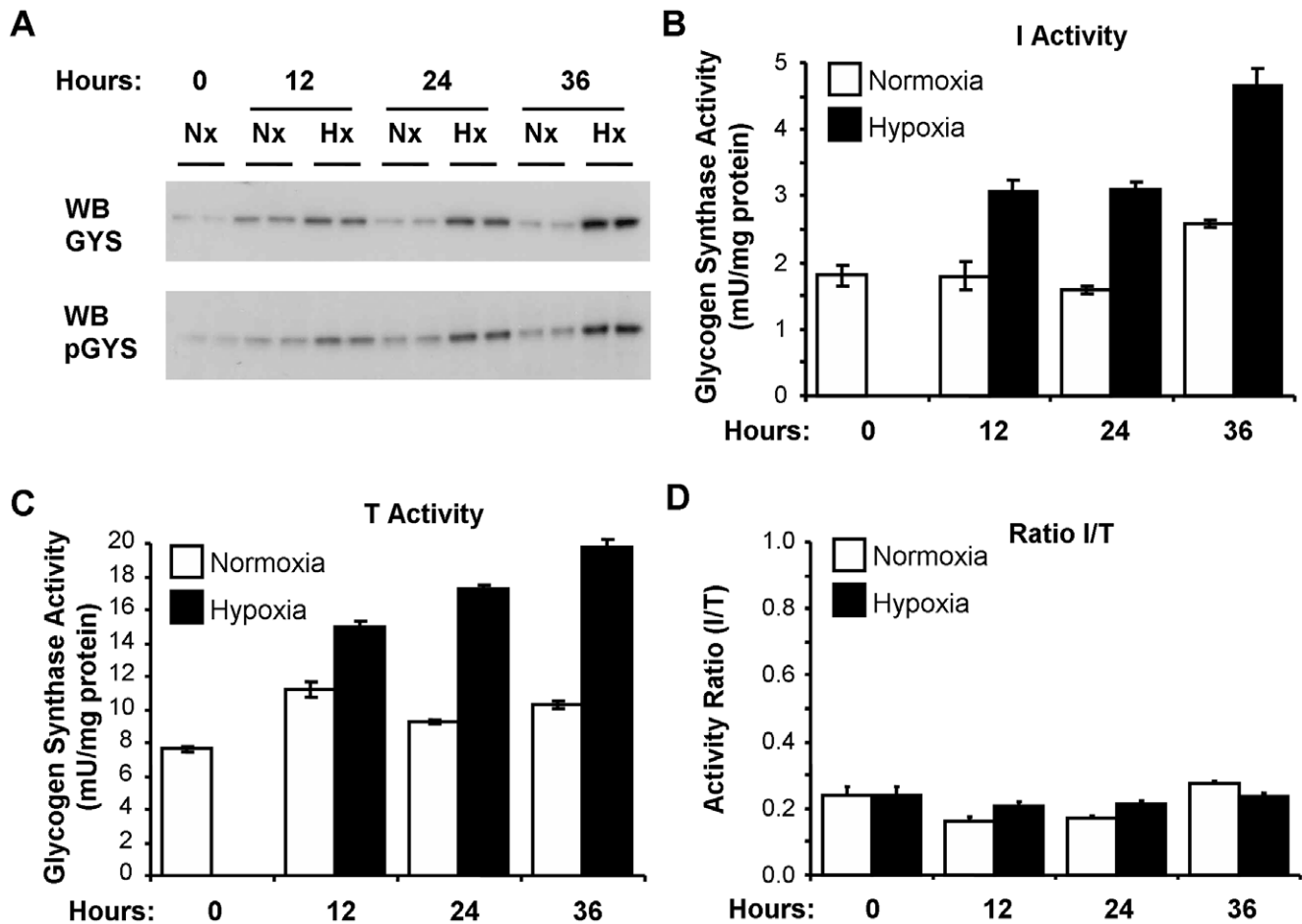


**Figure 3. Identification of a functional HRE in GYS1 genomic region.** (A) Schematic representation of human GYS1 genomic region from -429 to +298 showing evolutionarily conserved regions. Adapted from the UCSC Genome Browser (<http://genome.ucsc.edu/>). The lines above the graph show the genomic regions cloned in the indicated reporter constructs (see text). The inset under the figure represents the sequence alignment of the HRE element from the indicated species. (B) c2c12 cells exposed to 21% (Nx) or 1% (Hx) oxygen for 12 hours and then fixed and processed for chromatin immunoprecipitation with anti-HIF1 $\alpha$  (HIF1 $\alpha$ ) antibodies or control IgG (IgG). (Input) lanes correspond to total sonicated genomic DNA. After immunoprecipitation, chromatin was amplified with primers specific for the HRE-containing genomic regions from GYS1. The HIF binding region within the Procollagen 4-hydroxylase  $\alpha$  (P4Ha) was included as a positive control [25]. (C, D) HeLa cells transfected with a reporter plasmid containing GYS1 genomic region (-210 to -419) upstream of a minimal promoter (C) or a plasmid containing the whole GYS promoter region (+128 to -419) driving the expression of a luciferase reporter gene (D). Where indicated (mutHRE) the consensus HRE sequence (ACGT) was mutated to TAGC. The graphs represent the corrected luciferase activity values of each construct over the luciferase activity obtained in normoxic cells transfected with empty plasmids (control). Data shown are the results of four independent experiments and their mean (bar and numbers on the right). doi:10.1371/journal.pone.0009644.g003

Importantly, gene transcription correlates with an increment of muscle glycogen synthase protein (figure 2B and 4A). Since glycogen synthase (from now on GS) activity is subjected to a complex regulation by allosteric modulators and phosphorylation by several kinases, protein levels do not necessarily correlate with activity. Therefore, we next asked whether the increment in GYS1 protein resulted in augmented GS activity. Figure 4B shows that GS activity increased in parallel with protein accumulation (figure 4A). Importantly, in the presence of the

allosteric activator glucose-6-phosphate, that drives conversion of GS to its fully active state, samples from hypoxic cells still displayed increased GS activity as compared to control samples (figure 4C). Thus, hypoxia increased the total amount of GS but did not alter the proportion of active/inactive conformations as shown by the I/T ratio (figure 4D). In agreement, hypoxia did not modify the fraction of phosphorylated GYS1 (figure 4A). These results strongly suggest that GYS1 induction by HIF leads to augmented GS activity.





**Figure 4. Hypoxia increases glycogen synthase activity.** c2c12 myoblasts were incubated at 21% (Nx) or 1% oxygen (Hx) for the indicated periods of time and processed to determine GYS1/pGYS1(phospho-Ser 641) levels (A) and glycogen synthase activity in the absence (B) or presence of 6.6 mM of the allosteric modulator glucose-6-phosphate (C). Graphs represent the mean of duplicated measures and error bars the range. (D) The graph represents the ratio of glycogen synthase activity in the absence (I) and presence (T) of glucose-6-phosphate. The experiment was repeated twice with similar results.

doi:10.1371/journal.pone.0009644.g004

### Hypoxia induces glycogen accumulation

In order to test the functional relevance of increased GS activity, we exposed c2c12 myotubes to hypoxia (1% oxygen) or normoxia (21% oxygen) and determined their glycogen content at different time points. As shown in figure 5A, hypoxia resulted in a significant increase in glycogen concentration. Significantly, as noted for GYS1 gene induction (figure 1), glycogen accumulation was not restricted to muscle cells as it was also observed in liver-derived cells (figure 5B). Collectively, these results indicate that the hypoxic induction of GYS1 correlates with a significant increase in glycogen reserves.

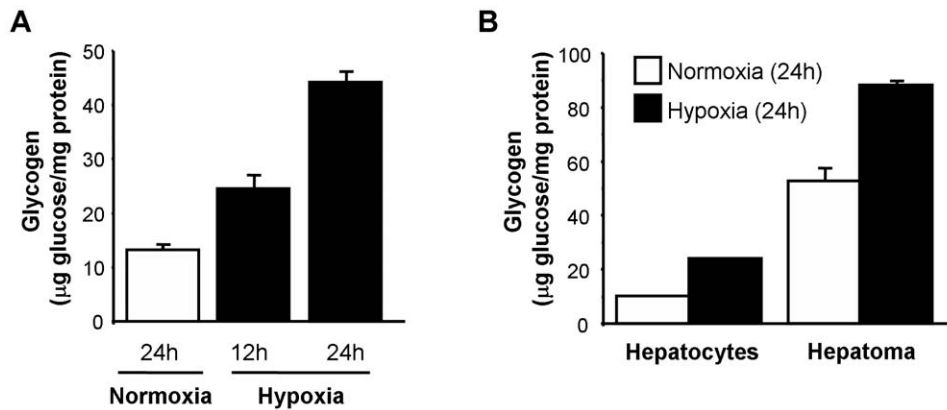
### Hypoxia-induced glycogen accumulation is HIF-dependent

As discussed before, HIF mediates most of the metabolic adaptations to hypoxia. Therefore, we next investigated the role of this transcription factor in glycogen accumulation. As shown in figure 6A, reduction of HIF activity by siRNA treatment attenuated the hypoxia-induced glycogen accumulation. In addition, the use of HepaC1/C4 cells provided further evidence of the role of HIF on the hypoxic accumulation of glycogen. While hypoxia induced glycogen accumulation in the HIF-competent HepaC1 cells, this effect was abrogated in the HIF-deficient

HepaC4 cells (figure 6B). These results demonstrate that HIF is necessary for the accumulation of glycogen induced by hypoxia.

### GYS1 mediates the hypoxic accumulation of glycogen

The results presented above demonstrate that hypoxia results in an increase in GYS1 expression and total glycogen synthase activity that correlates with glycogen accumulation. In order to establish a functional link between these phenomena, we investigated the effect of GYS1 interference on hypoxia-induced glycogen accumulation. Treatment of c2c12 with siRNA against GYS1 resulted in a reduction of both GYS1 mRNA and protein (figures 7A and 7B) that negatively affected glycogen accumulation induced by hypoxia as compared to control cells transfected with scramble siRNA (figure 7C). Since GS activity is an absolute requirement for glycogen synthesis, we next investigated the effect of knocking down GYS1 in the hepatoma HepaC1 cells, which also express the liver isoform GYS2. To this end, we infected HepaC1 cells with a retrovirus encoding for a short hairpin RNA directed to GYS1 (shGYS1) or an irrelevant shRNA (control), and selected pools of cells that stably integrated the construct. Expression of shGYS1 reduced GYS1 mRNA expression to  $12 \pm 7.7\%$  of control cells expressing a control shRNA (figure 7D) without affecting GYS2 expression (GYS2 mRNA level was



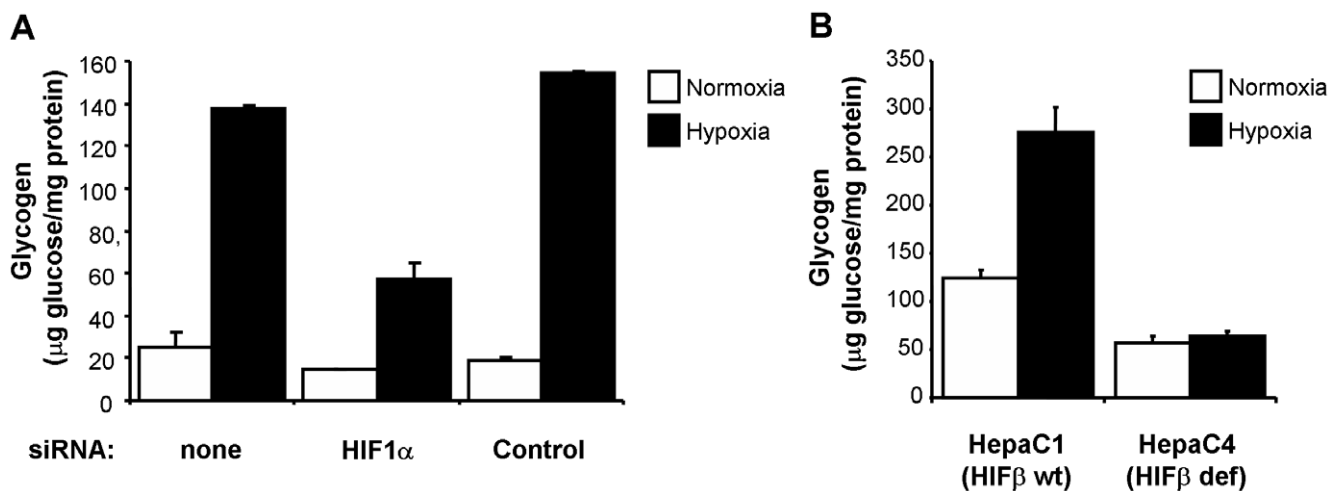
**Figure 5. Hypoxia induces glycogen accumulation.** C2c12 myotubes (A), immortalized mouse hepatocytes (B) and mouse hepatoma Hepa C1 cells (B), were exposed to normoxia or hypoxia (1% Oxygen) for the indicated periods of time and glycogen content was determined. Graphs represent average measures of glycogen content in two independent samples and error bars the range. Data shown are representative of at least five (A) or two (B) independent experiments.  
doi:10.1371/journal.pone.0009644.g005

67 $\pm$ 16% of that in control cells). In agreement with the experiments in c2c12 cells, knockdown of GYS1 prevented the hypoxia-induced accumulation of glycogen (figure 7E). Finally, we asked whether GYS1 induction was sufficient to promote glycogen accumulation. To this end, we determined the glycogen content of c2c12 cells infected with an adenovirus encoding for GYS1. As shown in figure 7F, forced expression of GYS1 led to the accumulation of glycogen to levels similar as those obtained under hypoxia. Altogether, these experiments indicate that GYS1 induction by HIF could explain the accumulation of glycogen observed under hypoxia.

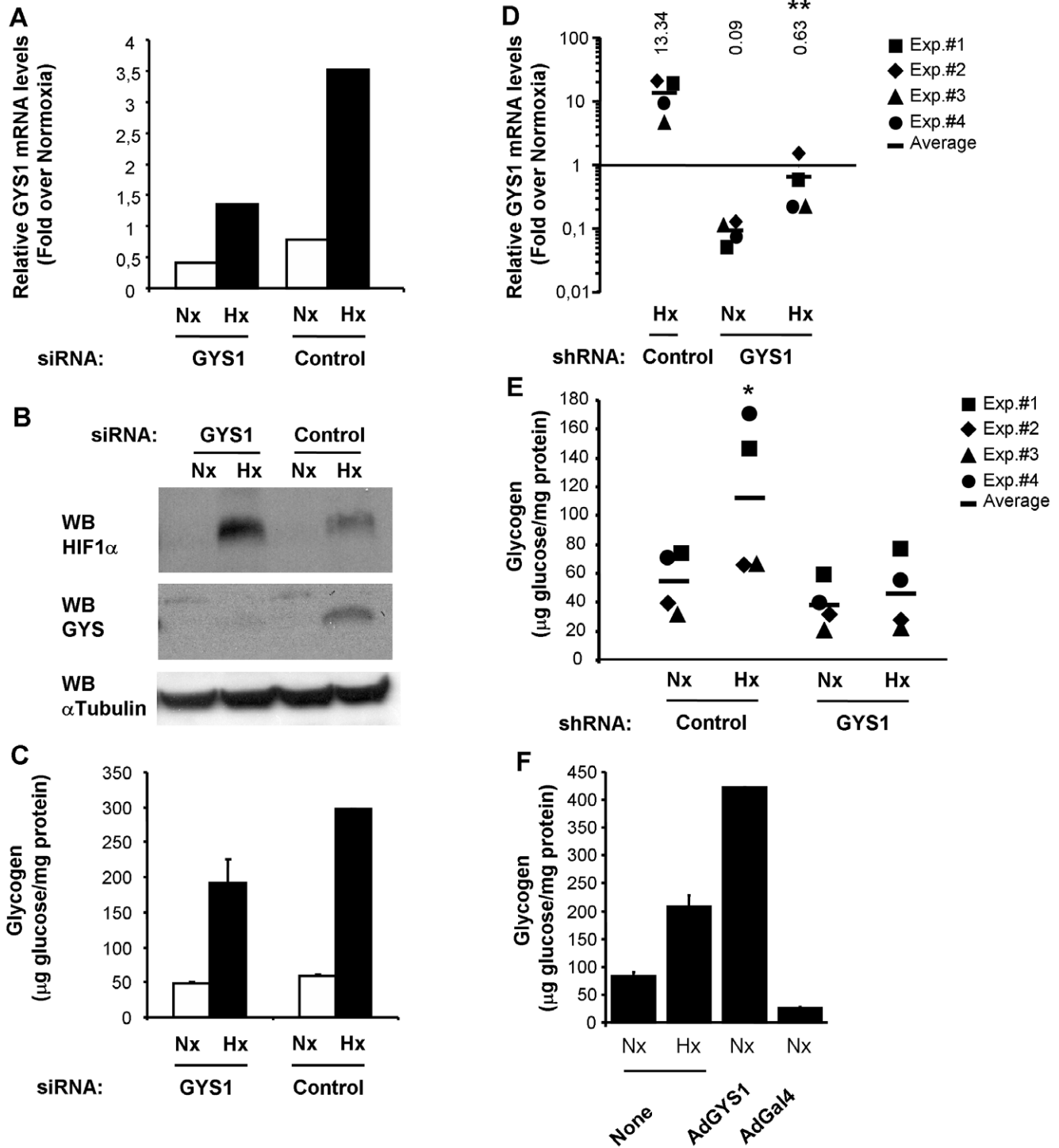
#### Hypoxia regulates glycogen metabolism at multiple levels to promote glycogen accumulation

The effect of hypoxia on glucose catabolism is mediated through the induction of almost all the enzymes of the glycolytic pathway in a coordinated fashion. Thus, we next investigated whether

hypoxia modulated other enzymes involved in glycogen synthesis. In fact, in addition to GYS1, our bioinformatics strategy (Ortiz et al. Nucl. Acids Res. in press) also identified UGP2 and GBE1 as potential HIF targets. Specifically we found two potential HREs in the UGP2 locus (located at positions -72 and +793, within the promoter region and first intron respectively) and one in the GBE1 locus (located within the first intron 8746 bp downstream of the transcription start site). UGP2 encodes for UTP:glucose-1-phosphate uridylyltransferase, an enzyme that synthesizes UDP-glucose, the substrate used by GS for glycogen synthesis. GBE1 encodes for 1,4- $\alpha$  glucan branching enzyme, the activity responsible for the generation of the highly ramified structure of glycogen. To investigate whether these genes were induced by hypoxia, we exposed c2c12 myoblasts and primary hepatocytes to 1% oxygen and determined the level of their mRNA by quantitative PCR. As shown in figure 8A, GBE1 mRNA, but not UGP2 mRNA, was strongly induced in c2c12 cells. In



**Figure 6. Hypoxia-driven glycogen accumulation is HIF-dependent.** (A) c2c12 myoblasts were transfected with siRNA against mouse HIF1 $\alpha$  (HIF1 $\alpha$ ), siRNA containing an irrelevant sequence (Control) or left untreated (none). After transfection cells were cultured for 12 hours under normoxia or 1% oxygen (Hypoxia) and processed to determine glycogen content. Graphs represent the average of two independent determinations and their range. Data shown are representative of three independent experiments. (B) Hepa C1 and Hepa C4 cells were exposed to 21% (Normoxia) or 1% oxygen (Hypoxia) for 12 hours and processed to determine their glycogen content. Graphs represent the mean of duplicated measures and error bars the range, results shown are representative of three independent experiments.  
doi:10.1371/journal.pone.0009644.g006



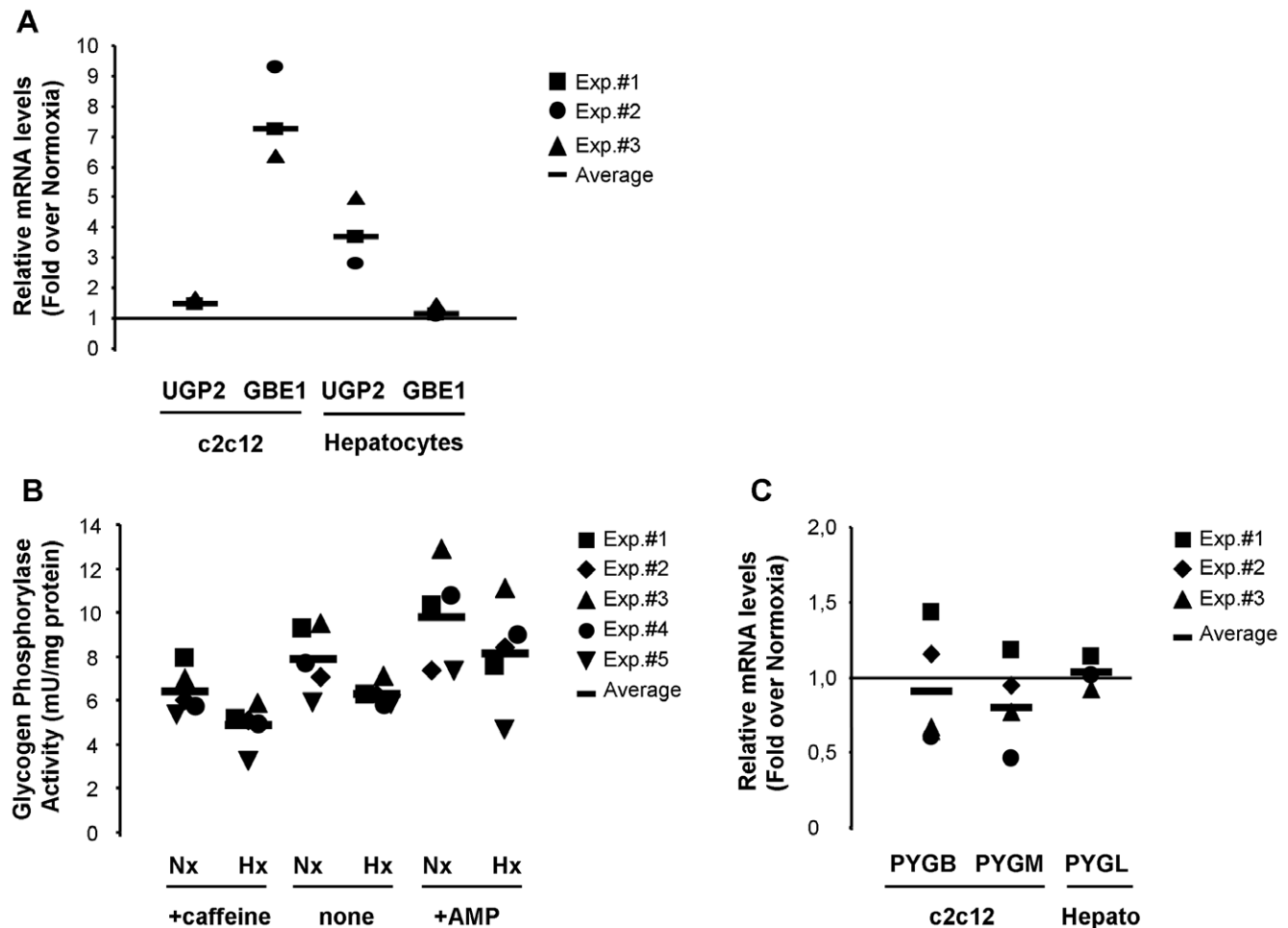
**Figure 7. GYS1 mediates glycogen accumulation during hypoxia.** c2c12 myoblasts were transfected with siRNA against mouse GYS1 (GYS1), siRNA containing an irrelevant sequence (Control). 24 hours after transfections cells were cultured for 12 hours under 21% (Nx) or 1% oxygen (Hx), then each sample was split and processed to determine GYS1 mRNA (A), protein levels (B) or glycogen content (C). The experiment was repeated twice with similar results. GYS1 mRNA levels (D) and glycogen content (E) were determined in Hepa C1 cells stably expressing a shRNA directed to GYS1 or a control shRNA. Results from four independent experiments are represented. \*,  $p < 0.05$ ; \*\*,  $p < 0.01$ ; as compared to samples expressing control shRNA and exposed to normoxia. (F) c2c12 myotubes were infected with AdGY1 or Ad $\beta$ -Gal. 24 hours after infection cells were processed for glycogen content determination. Parallel cultures were exposed to normoxia or hypoxia for 12 hours and processed for glycogen determination. Graphs represent the mean of duplicated measures and error bars the range, the experiment was repeated twice with similar results. doi:10.1371/journal.pone.0009644.g007

agreement with our data, a recent work demonstrated that GBE1 is induced by nickel treatment in a HIF-dependent manner [29]. On the other hand, UGP2 mRNA was upregulated by hypoxia in hepatocytes while GBE1 remained unaffected (figure 8A). The dissimilar behaviour of these genes in each cell type is probably a consequence of metabolic differences between muscle and liver. These differences notwithstanding, our results demonstrate that hypoxia induces the expression of all the enzymes required for glycogen synthesis. Finally, we also investigated whether, in addition to increased synthesis, a reduction of glycogen degradation could contribute to the accumulation of glycogen observed during hypoxia. In agreement with this possibility, we found a modest reduction of glycogen phosphorylase (GP) activity in cells exposed to hypoxia (figure 8B). In addition, this effect was observed even when the activity was assayed in the presence of the allosteric modulators caffeine and AMP that inhibit and activate GP, respectively (figure 8B). The effect of hypoxia on GP activity was reproducible and similar in magnitude to that induced by

caffeine and AMP, but differences did not reach statistical significance. Importantly, and in sharp contrast with the GS activity data, the reduction in GP activity did not correlate with altered gene expression, since none of the PYG isoforms encoding for glycogen phosphorylase were affected by hypoxia (figure 8C). Altogether, these results indicate that hypoxia leads to glycogen accumulation by affecting different reactions involved in its metabolism.

### Hypoxia-induced glycogen accumulation promotes cell survival

Cellular responses to hypoxia promote adaptation to low oxygen so that cells and tissues are better equipped to confront future oxygen restrictions. Therefore, we speculated that glycogen accumulation could contribute to the cellular adaptation to hypoxia by ensuring glucose supply for anaerobic glycolysis. To test this possibility, we first investigated the effect of hypoxia pre-treatment on the survival of cells deprived of



**Figure 8. Hypoxia affects glycogen metabolism at multiple levels.** (A) c2c12 myoblast and primary mouse hepatocytes were exposed to normoxia or hypoxia for 12 hours and the level of UGP2 and GBE1 mRNA was determined by quantitative PCR. The amount of each mRNA in samples was normalized to the content of  $\beta$ -actin mRNA in the same sample. Data shown represent the fold values of hypoxic over normoxic mRNA levels normalized to the value of 1 (horizontal line) in three independent experiments. (B) c2c12 myoblasts were exposed to normoxia (Nx) or hypoxia (Hx) for 24 hours and the activity of glycogen phosphorylase was assayed in untreated cell lysates, or in the presence of cAMP or caffeine. The results from five independent experiments and their mean are shown. (C) c2c12 myoblasts and primary mouse hepatocytes (Hepato) were exposed to normoxia or hypoxia for 12 hours and the level of PYGM, PYGB and PYGL mRNA was determined by quantitative PCR. The amount of each mRNA in samples was normalized to the content of  $\beta$ -actin mRNA in the same sample. Data shown represent the fold values of hypoxic over normoxic mRNA levels normalized to the value of 1 (horizontal line) in three independent experiments. doi:10.1371/journal.pone.0009644.g008

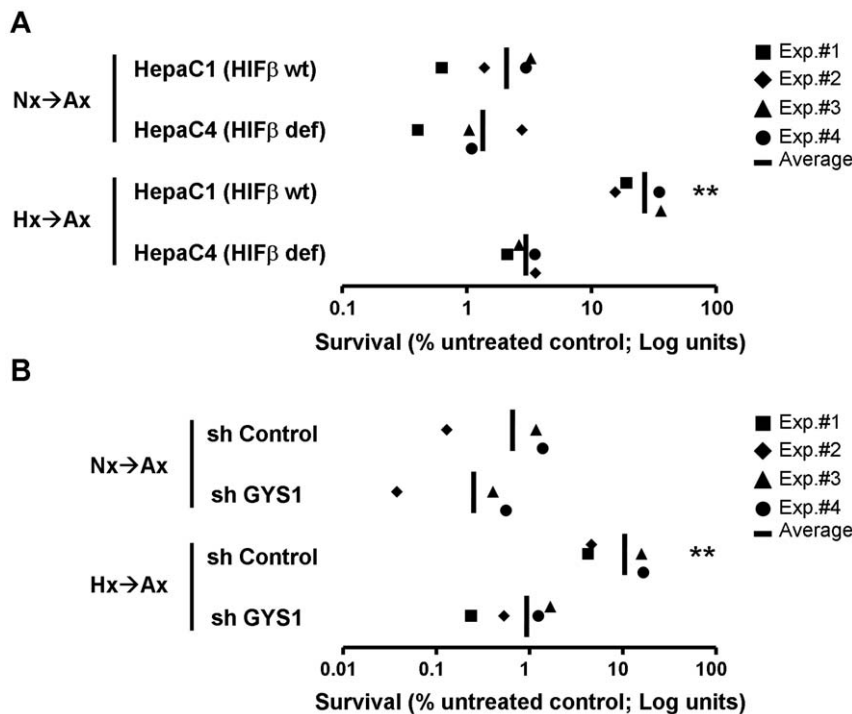
glucose and exposed to severe hypoxia. As shown in figure 9A, HepaC1 cells die within 12 hours after exposure to anoxia in the absence of glucose (Nx→Ax). Importantly, exposure of cells to hypoxia (1% oxygen for 24 h), prior to the anoxic treatment in the absence of glucose (Hx→Ax), resulted in increased survival (figure 9A). Thus, in this experimental setting, hypoxia triggers a response that improves cell survival upon subsequent hypoxic challenges (hypoxic preconditioning). Moreover, as observed in other systems, hypoxic preconditioning was dependent on HIF activity since it was not observed in HIF deficient HepaC4 cells (figure 9A). Next, we asked whether glycogen accumulation could be part of the HIF-dependent responses that mediate this preconditioning. To answer this question, we prevented hypoxia-induced glycogen accumulation in HepaC1 cells by expression of a short-hairpin RNA against GYS1, shGYS1 (figure 7D and E) and determined cell viability after exposure to anoxia in the absence of glucose. In agreement with a role for GYS1 and glycogen accumulation in hypoxic preconditioning, GYS1 repression resulted in decreased viability (figure 9B). Thus, these results suggest that glycogen accumulation plays a role in the adaptive response to hypoxia.

## Discussion

The data presented herein demonstrates that GYS1 is a novel hypoxia-inducible gene. The induction of GYS1 by hypoxia requires HIF-activity and an HRE motif located on GYS1 promoter. Consequently, GYS1 should be included among the growing cohort of HIF-target genes involved in glucose metabolism. In this regard, it is interesting to note that, similar to other HIF-regulated glycolytic genes [30], GYS1 promoter is induced by the HIF1 $\alpha$ , but not the HIF2 $\alpha$  isoform (Figure S2).

Importantly, the induction of GYS1 by HIF results in increased total GS activity during hypoxia without altering the I/T ratio. Thus, our results indicate that the increased GS activity is a consequence of the incremented amount of enzyme, rather than of a shift in its activation state. This result is significant since it underlines the importance of GS transcription in the regulation of glycogen synthesis, as opposed to the well-characterized posttranslational modifications of GS in response to hormones and metabolites. In addition, the effect of hypoxia is highly specific, as it affects GYS1, but not GYS2, gene transcription. To our knowledge, this is the first report of transcriptional regulation of GYS1 in response to environmental cues and the first report demonstrating the regulation of glycogen synthesis at the gene expression level. Moreover, recent findings demonstrate that the blockage of GYS1 degradation, caused by laforin or malin mutations, leads to increased glycogen levels in myoclonus epilepsy [31]. Hence, alteration of GS level by modifications of its transcription and/or degradation rate constitutes a novel form of regulation of the glycogen metabolism.

Importantly, we found that, in addition to GYS1, hypoxia also induced the expression of UGP2 and GBE1, genes that encode for the rest of the enzymatic activities required for glycogen synthesis. The regulation of GBE1 was previously described [29], but, to our knowledge, this is the first report suggesting UGP2 regulation by hypoxia. In agreement with our results, analysis of public gene expression datasets (Gene Expression Omnibus at NCBI, <http://www.ncbi.nlm.nih.gov>) reveals upregulation of UGP2 and GBE1 in several cell types exposed to hypoxia (Ortiz et al. Nucl. Acids Res. *in press*). Further work is required to address the role of HIF in their induction by hypoxia and to determine the functionality of the identified HREs. In spite of the central role of GS as the rate limiting step in glycogen accumulation, the induction of these



**Figure 9. GYS1 contributes to hypoxic preconditioning.** Hepa C1 and Hepa C4 cells (A) or Hepa C1 cells stably expressing a shRNA directed to GYS1 or a control shRNA (B) were exposed to 21% (Nx→Ax) or 1% (Hx→Ax) for 24 hours in complete media and then cultured in a glucose-free serum-free media and anoxia for 12 additional hours. Cell viability was determined as indicated in Materials and Methods. The results from four independent experiments and their mean are shown. doi:10.1371/journal.pone.0009644.g009

other genes surely contributes to an increased flow through this biosynthetic pathway. Additional experiments would be needed to address the relative contribution of these enzymes to hypoxic glycogen accumulation. On the other hand, our data suggest that hypoxia does not only regulate glycogen synthesis but also represses its degradation. This finding is in agreement with previous observations suggesting downregulation of GP activity during prolonged hypoxia [23]. Interestingly, the reduction in GP activity was not accompanied by decreased expression of the genes coding for the enzymes bearing this activity. However, GP activity from hypoxic cells was reduced even in the presence of the allosteric activator AMP, suggesting lower protein levels. In agreement with this possibility, preliminary experiments show a reduction in protein levels of the muscle isoform of GP in hypoxic c2c12 cells (data not shown). Thus it is plausible that hypoxia affects GP translation rate and/or protein stability.

Collectively, these results indicate that hypoxia induces a coordinated response that affects glycogen metabolism at multiple levels, in a similar fashion to the paradigmatic hypoxic regulation of glycolysis. Finally, we cannot rule out that other hypoxia-triggered pathways play a role in the accumulation of glycogen. In this regard, it is likely that the upregulation of the glucose transporter SLC2A1 (Glut-1) by HIF contributes to the accumulation of glycogen by increasing the intracellular concentration of glucose.

Since anaerobic glycolysis is characterized by increased glucose demand, the induction of glycogen synthesis during hypoxia might seem paradoxical. However, as the induction of glycogen synthesis requires HIF-dependent gene expression, it will be effective during chronic hypoxia and/or subsequent hypoxic episodes. Therefore, the induction of glycogen synthesis during prolonged hypoxia might serve to replenish glycogen stores that could have been mobilized during early acute hypoxia. In agreement, it was previously proposed [22,23] that muscle glycogen supercompensation, the accumulation of glycogen above basal levels observed after depletion of reserves during intense muscle exercise, was a consequence of the hypoxia-induced glycogen accumulation. Additionally, the induction of glycogen synthesis during chronic hypoxia results in larger cellular reserves. Our cell viability results suggest that hypoxic glycogen accumulation moderately improves cell survival following a hypoxic/ischemic insult. It should be noted however, that our assay could reflect a higher metabolic rate rather than enhanced viability. At any rate, this result implies that hypoxia-exposed cells are better equipped to respond to subsequent hypoxic challenges due to, at least in part, their higher glycogen content. In accord, hypoxia tolerant animals present large glycogen stores that are critical to ensure glucose supply during oxygen restriction [1]. Thus, we propose that the induction of GYS1 by HIF constitutes a cellular response involved in long-term adaptation to hypoxia that prepare cells to better cope with future oxygen restrictions.

Understanding the molecular mechanism of this novel response could be beneficial, not only in elucidating the physiology of the adaptations to hypoxia, but also in rationally designing novel therapeutic strategies to treat pathologies that course with hypoxia. In this regard, excessive accumulation of glycogen causes several pathologies that affect heart and central nervous system physiology. Hence, it is possible that GYS1 induction, while being an adaptive response that protects cells, could lead to an excessive glycogen accumulation in heart or brain that could contribute to damage in these organs following ischemic insults. In agreement with this possibility, it has been recently reported that chronic intermittent hypoxia leads to liver injury due, at least in part, to glycogen accumulation in hepatocytes [32]. On the other hand, our results showing that hypoxia-induced glycogen accumulation

protects cells from subsequent ischemia could be relevant to tumor biology. As a consequence of aberrant vasculature, tumors are exposed to intermittent hypoxia that could induce glycogen accumulation. If this response is present, it could contribute to the resistance to fluctuations in blood supply that is commonly observed in tumors. Further work is required to test this hypothesis and to investigate the potential role of GYS1 as a target for therapeutic intervention in cancer.

In summary, the results presented herein reveal a novel metabolic adaptation of cells to function under reduced oxygen tensions. Our results further emphasize the central role of the metabolic adaptation in the response to hypoxia and the anticipatory nature of this response, as it seeks to prevent or reduce the effects of a more severe hypoxia.

## Materials and Methods

**Cell Culture and Reagents**— c2c12 myoblasts were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum. To induce myogenic differentiation (c2c12 myotubes), cells were allowed to reach confluence and were then cultured in DMEM supplemented with 2% horse serum. Neonatal hepatocytes were kindly provided by Angela Martínez Valverde (IIB, Madrid). HepaC1 and HepaC4 cells were maintained in  $\alpha$ -modified Eagle's medium (Biochrom) supplemented with 10% fetal bovine serum. In all cases, culture medium was supplemented with 100 units/ml penicillin and 100  $\mu$ g/ml streptomycin. Cells were grown at 37°C and 5% CO<sub>2</sub> in a humidified incubator. For hypoxia treatments, cells were grown at 37°C in sealed chambers (Billups-Rothenberg) flushed with a 1% O<sub>2</sub>, 5% CO<sub>2</sub>, 94% N<sub>2</sub> gas mixture or in a Whitley hypoxystation (don Whitley Scientific, UK) set at the indicated oxygen concentration. Dimethylxalylglycine (DMOG, Frontier Scientific, CA, USA) was added to the indicated cultures at a 250  $\mu$ M final concentration.

**Plasmid Construction**— Human genomic DNA extracted from human cervical carcinoma (HeLa) cells was used as template for PCR amplification of GYS1 promoter regions. PCR products were first cloned into pCR2.1-TOPO (Invitrogen) and subsequently subcloned for reporter assays into the KpnI/XhoI restriction sites of pGL3-Basic (Invitrogen) or the BamHI site of prolactin- plasmid [33], in both possible orientations. The identity of all constructs was verified by sequencing. Primer sequences are available from the authors upon request.

**Reporter assays**— Reporter assays were performed using the human cervical-carcinoma cell line HeLa. Cells were seeded on six-well plates (3.5\*10<sup>4</sup> cells/well) 6 h prior to transfection. A 9  $\mu$ g DNA mixture containing 3  $\mu$ g of the indicated reporter construct or empty plasmid and 0.1  $\mu$ g of a plasmid encoding for *Renilla* (sea pansy) luciferase under the control of a null promoter (Promega, Madison, WI, U.S.A.) was used for transfection using the calcium phosphate method. 16 h after transfection, cells were washed, replated in 24-well plates, and incubated in normoxia, in the presence of DMOG or in hypoxia for an additional 16 hours. After treatments cells were lysed and the firefly and *Renilla* luciferase activities of the lysate were determined using a dual-luciferase system (Promega, Madison, WI, U.S.A.). The firefly luciferase activity was normalized to that of *Renilla* luciferase.

**Treatment with Recombinant Adenoviruses**— AdCMV-GFP-MGS and AdCMV-galactosidase were described previously [34]. For infections cells were incubated for 2 h with an appropriate amount of adenovirus in OptiMEM. Infection medium was then replaced with DMEM, and cells were incubated for 48 h at 37°C in a humidified incubator.

**RNA interference-** We used the following commercial siRNAs: HIF1 $\alpha$  siRNA (Santa Cruz, pre-designed siRNA#16708), GYS1 siRNA (Qiagen, SI01060843), and All Stars siRNA (Qiagen, catalog # 1027281) as a negative control. A concentration of 30 nM of each siRNA was transfected into c2c12 cells using Lipofectamine 2000 (Invitrogen), following the manufacturer's instructions. The efficiency of HIF1 $\alpha$  or GYS1 knockdown was determined by real time quantitative PCR analysis. Stable attenuation of GYS1 expression was achieved by selection of pooled cells infected with lentivirus expressing a short hairpin RNA directed to mouse GYS1. The lentiviral constructs were a kind gift from Carles Martinez-Pons (IRB, Barcelona).

**Metabolite determinations-** Glycogen determination was performed as previously described [35]. Briefly, cells were scraped with 30% (w/v) KOH; the extract was then boiled for 15 min and the resulting solution was spotted on chromatography paper 31 ET (Whatman). Glycogen was precipitated by immersing the papers in ice-cold 66% (v/v) ethanol. After two washes in ethanol, the papers were air-dried and incubated with amyloglucosidase (Sigma). The resulting glucose was measured with a Gluco-quant kit (ABX entra). The amount of glycogen is represented as the amount of released glucose per mg of total protein. L-Lactate levels in the incubation medium were measured as described [36].

**GS and GP activity determination-** To measure GS and GP activities, frozen plates were scraped with 300  $\mu$ l of homogenization buffer consisting of 10 mM Tris/HCl (pH 7.0), 150 mM KF, 15 mM EDTA, 600 mM sucrose, 15 mM 2-mercaptoethanol, 10  $\mu$ g/ml leupeptin, 1 mM benzamidin and 1 mM PMSF, and the collected cells were subsequently sonicated. The resulting homogenates were used for the determination of enzymic activities. Protein concentration was measured by the Bio-Rad protein assay. GP activity was determined by measuring the incorporation of [U-<sup>14</sup>C]glucose 1-phosphate into glycogen in the absence or presence of AMP (5 mM) as describe previously [37]. The GS activity ratio was determined by measuring the incorporation of [U-<sup>14</sup>C]UDP-glucose into glycogen in the absence or presence of 6.6 mM glucose 6-phosphate, as described previously [38]. The activity measured in the absence of Glc-6-P represents the active form of the enzyme (I or a form), whereas that measured in the presence of 6.6 mM Glc-6-P represents total GS activity.

**Western Blot**—Immediately after treatments, cells were washed with ice-cold phosphate-buffered saline and harvested in 70–200  $\mu$ l of 1x Laemmli sample buffer (2%SDS, 10% glycerol, 5% 2-mercaptoethanol, 0.002% bromphenol blue and 0.125 M Tris HCl pH 6.8). Lysates were sonicated for 4 s, centrifuged at 4°C for 2 min at 14,000 $\times$ g, and resolved on 8–10% SDS-polyacrylamide gels. Proteins were then transferred to nitrocellulose membranes (Bio-Rad), blocked with 5% nonfat dry milk in TBS-T (50 mM Tris, pH 7.6, 150 mM NaCl, 0.1% Tween 20), and incubated overnight at 4°C with the indicated antibodies (HIF1 $\alpha$  was from R&D systems, ref. MAB1536 and GYS1 was from Cell Signalling, ref 3893). Immunoreactive bands were visualized with the Amersham Biosciences ECL system.

**ChIP (chromatin immunoprecipitation)-** For ChIP assays, c2c12 cells were grown on 10 cm plates until they reached 85% confluence, at which point they were exposed to hypoxia (1% oxygen) or left under normoxic conditions for a further 5 h. Subsequently, cells were fixed with 1% (v/v) formaldehyde (final concentration) for 12 min at 37°C. Crosslinking was stopped by the addition of 0.125 M glycine (final concentration). The cells were washed with cold PBS and then lysed by scraping in 1 ml of lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris/HCl, pH 8.1, and a protease inhibitor cocktail, Roche). Cell lysates were incubated on ice for

10 min and then sonicated to shear the DNA under conditions established to ensure that the DNA fragments were between 200 and 1500 bp. After the removal of the insoluble material by centrifugation, 30  $\mu$ l of each sample was removed and stored (input), while the rest was diluted in immunoprecipitation buffer (1% Triton X-100, 2 mM EDTA, 150 mM NaCl and 20 mM Tris/HCl, pH 8.1). The lysates were precleared with preimmune serum and 200  $\mu$ g of a Salmon Sperm DNA/Protein A agarose 50% slurry (Upstate Biotechnology, Lake Placid, NY, U.S.A.) for 1 h at 4°C. The samples were then immunoprecipitated twice, initially with whole rabbit serum for 6 h (IgG control) and then overnight at 4°C with a polyclonal anti-HIF1 $\alpha$  antiserum (Abcam, ab2185). Immunocomplexes were recovered by the addition of 400  $\mu$ g of Salmon Sperm DNA/Protein A agarose 50% slurry to the samples that were then sequentially washed for 15 min in TSE I (0.1%SDS, 1%Triton X-100, 2 mM EDTA, 20 mM Tris/HCl, pH 8.1, and 150 mM NaCl), TSE II (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris/HCl, pH 8.1, and 500 mM NaCl) and buffer III (0.25 M LiCl, 1% Nonidet P40, 1% deoxycholate, 1 mM EDTA and 10 mM Tris/HCl, pH 8.1). Finally, the complexes were washed twice with TE buffer (10 mM Tris, pH 8.0, and 1 mM EDTA) and extracted twice with a buffer containing 1% SDS and 0.1 M NaHCO<sub>3</sub>. The eluates were pooled, and cross-linking was reversed by the addition of 200 mM NaCl (final concentration) and overnight incubation at 65°C. The proteins were removed by the addition of proteinase K (30  $\mu$ g/sample) for 2 h at 42°C, and the DNA was extracted using Qiagen PCR extraction kit before eluting in 50  $\mu$ l of water. Immunoprecipitated DNA was amplified by PCR using the primers indicated:

GYS1F:TCGGATCCGGTACCGCTTTACGGAAACGAGTG  
 GYS1R:TACTCGAGGGTACCGAACTCCTGGGTCCATTC  
 P4HaF: ATCAAGGAGGCAAACCTGAACAG  
 P4HR: ACTCGGAGCGGCTACTTCCTA

The PCR products were resolved by gel electrophoresis and visualized by ethidium bromide staining.

**RNA Extraction and Quantitative PCR**—Total RNA was extracted and purified with the RNeasy Mini Kit (Qiagen). 1  $\mu$ g of total RNA from each sample was retrotranscribed to cDNA (Improm-II reverse transcriptase; Promega). 1–3  $\mu$ l of cDNA samples were used as template for amplification reactions carried out with the LC Fast Start DNA master SYBR Green I kit (Roche Applied Science) following the manufacturer's instructions. PCR amplifications were carried out in a Light Cycler System (Roche Applied Science), and data were analyzed with LightCycler software 3 version 3.5.28 (Idaho Technology Inc.). For each sample, duplicate determinations were made, and the gene expression determined by the  $\Delta\Delta$ Ct method using  $\beta$ -actin as reference gene. The primer used in this study are (5'-3'): GYS1, forward (TTCTACAACAACCTGGAG) and reverse (CTGAGCAGATAGTTGAGC); GYS2, forward (TTACCAGCATGCCAGACAC) and reverse (AGAAGGTGGTACTGAGG); HIF 1 $\alpha$ , forward (GTTTACTAAAGGCAAGTCACC) and reverse (TTCTGTTTGTGAAGGGAG);  $\beta$ -actin forward (CCCAGAGCAAGAGAGG) and reverse (GTCCA-GACGCAGGATG). UGP2, forward (CAGAGACCTCCAGAA-GATTTCG) and reverse (GTTCAACACAGAAGATATGTTAT-CAGG); GBE1, forward (GCTCGGTGGAGAAGGCTAT) and reverse (CTTGGGAAGTCCAACCATTTC).

**Cell viability assay.**—HepaC1, HepaC4, HepaC1-Ad1 (stably expressing a control shRNA) and HepaC1-Ad5 cells (stably expressing a shRNA directed to mouse GYS1) were seeded in 96-well plates at a density of 5 $\times$ 10<sup>3</sup>/well in 100  $\mu$ l of complete medium. Then the cells were incubated in normoxia or in hypoxia for 24 h. After treatments culture media was removed, cells were

washed, changed to glucose and serum free medium, and incubated in normoxia or in anoxia for additional 16 hours. Cell viability was determined with the CellTiter-Glo cell viability assay (Promega, WI). Luciferase activity was detected in a microplate luminometer reader (E6521; Promega, WI).

**Statistical analysis of data**— Experimental data were analyzed with the Prism™ GraphPad (version 4.01) software. Data were analyzed by the analysis of variance test (ANOVA) followed by the Tukey's or Dunnett's multiple comparison test. Significant differences with control values (indicated in each figure legend) are shown by asterisks.

## Supporting Information

**Figure S1** GYS1 induction in different cell types. The indicated cell lines were exposed to normoxia or hypoxia for 6 hours and the level of GYS1 mRNA was determined by quantitative PCR. The amount of each mRNA in samples was normalized to the content of beta-actin mRNA in the same sample. Data shown represent the fold values of hypoxic over normoxic mRNA levels normalized to the value of 1 (horizontal line). The experiment was repeated twice with similar results.

Found at: doi:10.1371/journal.pone.0009644.s001 (1.24 MB TIF)

**Figure S2** Promoter activation by HIF1a or HIF2a overexpression. HeLa cells transfected with a reporter plasmid containing

BNIP3 (A), CITED2 (B), VEGFA (C) or GYS1 (D) promoters upstream a luciferase reporter gene alone (none) or in combination with constructs encoding for HIF1a or HIF2a. The graphs represent the corrected luciferase activity values of each construct over the luciferase activity obtained in normoxic cells transfected with empty plasmids. Data shown are the average results of three independent experiments and error bars the standard deviation. Found at: doi:10.1371/journal.pone.0009644.s002 (2.33 MB TIF)

## Acknowledgments

We thank Angela Martínez Valverde for primary mouse hepatocytes and Juan Miguel Redondo, Arantazu Alfranca, Rocío Aguilar and Éva Siles Rivas for help with in vivo hypoxia experiments. Reporter plasmids containing CITED2 and BNIP3 promoters were a kind gift from S. Bhattacharya and Olga Aprelikova, and Richard Bruick, respectively. We thank Julian Aragones for critical reading of this manuscript and his valuable suggestions.

## Author Contributions

Conceived and designed the experiments: LdP. Performed the experiments: NP DV DC MGR AOB SV YC DSM. Analyzed the data: NP DV AOB AO LdP. Contributed reagents/materials/analysis tools: MLGB MOLJG LdP. Wrote the paper: LdP.

## References

- Ramirez J-M, Folkow LP, Blix AS (2007) Hypoxia Tolerance in Mammals and Birds: From the Wilderness to the Clinic. *Annual Review of Physiology* 69: 113–143.
- Bartrons R, Caro J (2007) Hypoxia, glucose metabolism and the Warburg's effect. *Journal of Bioenergetics and Biomembranes* 39: 223–229.
- Kim J-w, Tchernyshyov I, Semenza GL, Dang CV (2006) HIF-1-mediated expression of pyruvate dehydrogenase kinase: A metabolic switch required for cellular adaptation to hypoxia. *Cell Metabolism* 3: 177–185.
- Papandreou I, Cairns RA, Fontana L, Lim AL, Denko NC (2006) HIF-1 mediates adaptation to hypoxia by actively downregulating mitochondrial oxygen consumption. *Cell Metabolism* 3: 187–197.
- Semenza G, Roth P, Fang H, Wang G (1994) Transcriptional regulation of genes encoding glycolytic enzymes by hypoxia-inducible factor 1. *J Biol Chem* 269: 23757–23763.
- Fukuda R, Zhang H, Kim JW, Shimoda L, Dang CV, et al. (2007) HIF-1 regulates cytochrome oxidase subunits to optimize efficiency of respiration in hypoxic cells. *Cell* 129: 111–122.
- Gu YZ, Moran SM, Hogenesch JB, Wartman L, Bradfield CA (1998) Molecular characterization and chromosomal localization of a third alpha-class hypoxia inducible factor subunit, HIF3alpha. *Gene Expr* 7: 205–213.
- Wang GL, Jiang BH, Rue EA, Semenza GL (1995) Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O<sub>2</sub> tension. *Proc Natl Acad Sci USA* 92: 5510–5514.
- Wiesener MS, Turley H, Allen WE, Willam C, Eckardt KU, et al. (1998) Induction of endothelial PAS domain protein-1 by hypoxia: characterization and comparison with hypoxia-inducible factor-1alpha. *Blood* 92: 2260–2268.
- Huang LE, Gu J, Schau M, Bunn HF (1998) Regulation of hypoxia-inducible factor 1alpha is mediated by an O<sub>2</sub>-dependent degradation domain via the ubiquitin-proteasome pathway. *Proc Natl Acad Sci U S A* 95: 7987–7992.
- Ivan M, Kondo K, Yang H, Kim W, Valiando J, et al. (2001) HIF1alpha targeted for VHL-mediated destruction by proline hydroxylation: implications for O<sub>2</sub> sensing. *Science* 292: 464–468.
- Jaakkola P, Mole DR, Tian YM, Wilson MI, Gielbert J, et al. (2001) Targeting of HIF-1alpha to the von Hippel-Lindau ubiquitylation complex by O<sub>2</sub>-regulated prolyl hydroxylation. *Science* 292: 468–472.
- Maxwell PH, Wiesener MS, Chang GW, Clifford SC, Vaux EC, et al. (1999) The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature* 399: 271–275.
- Yu F, White SB, Zhao Q, Lee FS (2001) HIF-1alpha binding to VHL is regulated by stimulus-sensitive proline hydroxylation. *Proc Natl Acad Sci U S A* 98: 9630–9635.
- Bruick RK, McKnight SL (2001) A conserved family of prolyl-4-hydroxylases that modify HIF. *Science* 294: 1337–1340.
- Epstein AC, Gleadow JM, McNeill LA, Hewitson KS, O'Rourke J, et al. (2001) C. elegans EGL-9 and Mammalian Homologs Define a Family of Dioxygenases that Regulate HIF by Prolyl Hydroxylation. *Cell* 107: 43–54.
- Ivan M, Haberberger T, Gervasi DC, Michelson KS, Gunzler V, et al. (2002) Biochemical purification and pharmacological inhibition of a mammalian prolyl hydroxylase acting on hypoxia-inducible factor. *Proc Natl Acad Sci U S A* 99: 13459–13464.
- Wenger RH, Stiehl DP, Camenisch G (2005) Integration of Oxygen Signaling at the Consensus HRE. *Sci STKE* 2005: re12.
- Aragones J, Schneider M, Van Geyte K, Fraisl P, Dresselaers T, et al. (2008) Deficiency or inhibition of oxygen sensor Phd1 induces hypoxia tolerance by reprogramming basal metabolism. *Nat Genet* 40: 170–180.
- Firth JD, Ebert BL, Ratcliffe PJ (1995) Hypoxic Regulation of Lactate Dehydrogenase A. *J Biol Chem* 270: 21021–21027.
- Ullah MS, Davies AJ, Halestrap AP (2006) The Plasma Membrane Lactate Transporter MCT4, but Not MCT1, Is Up-regulated by Hypoxia through a HIF-1alpha-dependent Mechanism. *J Biol Chem* 281: 9030–9037.
- Mamedova LK, Shneyvays V, Katz A, Shainberg A (2003) Mechanism of glycogen supercompensation in rat skeletal muscle cultures. *Molecular and Cellular Biochemistry* 250: 11–19.
- Vigoda A, Mamedova LK, Shneyvays V, Katz A, Shainberg A (2003) Glycogen metabolism in rat heart muscle cultures after hypoxia. *Molecular and Cellular Biochemistry* 254: 311–318.
- Sridharan V, Guichard J, Bailey RM, Kasiganesan H, Beeson C, et al. (2007) The prolyl hydroxylase oxygen-sensing pathway is cytoprotective and allows maintenance of mitochondrial membrane potential during metabolic inhibition. *AJP - Cell Physiology* 292: C719–728.
- Pescador N, Cuevas Y, Naranjo S, Alcaide M, Villar D, et al. (2005) Regulation of the egl nine homologue 3 (egl3/phd3) gene: Identification of a functional hypoxia-responsive element. *Biochemical Journal* 390: 189–197.
- Wood SM, Gleadow JM, Pugh CW, Hankinson O, Ratcliffe PJ (1996) The role of the aryl hydrocarbon receptor nuclear translocator (ARNT) in hypoxic induction of gene expression. *Studies in ARNT-deficient cells. J Biol Chem* 271: 15117–15123.
- Keiko N-T, Akira K, Kazuhiro S, Yoshiaki F-K (1997) A point Mutation Responsible for Defective Function of the Aryl-Hydrocarbon-Receptor Nuclear Translocator in Mutant Hepa-1c1c7 Cells. *European Journal of Biochemistry* 246: 486–495.
- Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, et al. (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Research* 15: 1034–1050.
- Zhao J, Chen H, Davidson T, Kluz T, Zhang Q, et al. (2004) Nickel-induced 1,4-[alpha]-glucan branching enzyme 1 up-regulation via the hypoxic signaling pathway. *Toxicology and Applied Pharmacology* 196: 404–409.
- Hu C-J, Wang L-Y, Chodosh LA, Keith B, Simon MC (2003) Differential Roles of Hypoxia-Inducible Factor 1alpha (HIF-1alpha) and HIF-2alpha in Hypoxic Gene Regulation. *Mol Cell Biol* 23: 9361–9374.
- Vilchez D, Ros S, Cifuentes D, Pujadas L, Valles J, et al. (2007) Mechanism suppressing glycogen synthesis in neurons and its demise in progressive myoclonus epilepsy. *Nat Neurosci* 10: 1407–1413.
- Savransky V, Nanayakkara A, Vivero A, Li J, Bevans S, et al. (2007) Chronic intermittent hypoxia predisposes to liver injury. *Hepatology* 45: 1007–1013.



33. Aragonés J, Jones DR, Martín S, San Juan MA, Alfranca A, et al. (2001) Evidence for the involvement of diacylglycerol kinase in the activation of hypoxia-inducible transcription factor 1 by low oxygen tension. *J Biol Chem* 276: 10548–10555.
34. Ros S, Garcia-Rocha M, Dominguez J, Ferrer JC, Guinovart JJ (2009) Control of Liver Glycogen Synthase Activity and Intracellular Distribution by Phosphorylation. *Journal of Biological Chemistry* 284: 6370–6378.
35. Chan TM, Exton JH (1976) A rapid method for the determination of glycogen content and radioactivity in small quantities of tissue or isolated hepatocytes. *Anal Biochem* 71: 96–105.
36. Gutmann I, Wahlefeld AW (1974) *Methods of Enzymatic Analysis*. New York: Academic Press Inc.
37. Gilboe DP, Larson KL, Nuttall FQ (1972) Radioactive method for the assay of glycogen phosphorylases. *Analytical Biochemistry* 47: 20–27.
38. Thomas JA, Schlender KK, Larner J (1968) A rapid filter paper assay for UDPglucose-glycogen glucosyltransferase, including an improved biosynthesis of UDP-14C-glucose. *Analytical Biochemistry* 25: 486–499.