Identification of non-coding genetic variants in samples from hypoxemic respiratory disease patients that affect the transcriptional response to hypoxia

Olga Roche[1,2,3], María Laura Deguiz[1,2], María Tiana[1,2,4], Clara Galiana-Ribote[1], Daniel Martinez-Alcazar[1], Carlos Rey-Serra[1], Beatriz Ranz-Ribeiro[1], Raquel Casitas[2,4,5], Raúl Galera[2,4,5], Isabel Fernández-Navarro[4,5], Silvia Sánchez-Cuéllar[4,6], Virginie Bernard[7], Julio Ancochea[4,6,8], Wyeth W. Wasserman[7], Francisco García-Rio[2,4,5,8], Benilde Jimenez[1,2,4] and Luis del Peso[1,2,4,*]

[1]Departamento de Bioquímica, Universidad Autónoma de Madrid (UAM) and Instituto de Investigaciones Biomédicas 'Alberto Sols' (CSIC-UAM), 28029 Madrid, Spain, [2]IdiPaz, Instituto de Investigación Sanitaria del Hospital Universitario La Paz, 28029 Madrid, Spain, [3]Unidad de Medicina Molecular, laboratorio de Oncología, CRIB. Universidad de Castilla-La Mancha, 02006, Albacete, Spain, [4]Centro de Investigación Biomédica en Red de Enfermedades Respiratorias (CIBERES), Instituto de Salud Carlos III, 28029 Madrid, Spain, [5]Hospital Universitario La Paz, IdiPAZ, Servicio de Neumología, 28029 Madrid, Spain, [6]Servicio de Neumología, Hospital Universitario de La Princesa, Instituto de Investigación Sanitaria del hospital de La Princesa, 28006 Madrid, Spain, [7]Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, Department of Medical Genetics, University of British Columbia Vancouver, British Columbia V5Z 4H4, Canada and [8]Departamento de Medicina, Universidad Autónoma de Madrid (UAM) 28029 Madrid, Spain

ABSTRACT

A wide range of diseases course with an unbalance between the consumption of oxygen by tissues and its supply. This situation triggers a transcriptional response, mediated by the hypoxia inducible factors (HIFs), that aims to restore oxygen homeostasis. Little is known about the inter-individual variation in this response and its role in the progression of disease. Herein, we sought to identify common genetic variants mapping to hypoxia response elements (HREs) and characterize their effect on transcription. To this end, we constructed a list of genome-wide HIF-binding regions from publicly available experimental datasets and studied the genetic variability in these regions by targeted re-sequencing of genomic samples from 96 chronic obstructive pulmonary disease and 144 obstructive sleep apnea patients. This study identified 14 frequent variants disrupting potential HREs. The analysis of the genomic regions containing these variants by means of reporter assays revealed that variants rs1009329, rs6593210 and rs150921338 impaired the transcriptional response to hypoxia. Finally, using genome editing we con-firmed the functional role of rs6593210 in the transcriptional regulation of EGFR. In summary, we found that inter-individual variability in non-coding regions affect the response to hypoxia and could potentially impact on the progression of pulmonary diseases.

INTRODUCTION

Hypoxia, defined as the imbalance between cellular oxygen demand and its supply, is a common and central feature in highly prevalent pathological conditions including respiratory, cardiovascular and inflammatory diseases as well as neoplasias ([1]). In response to a decrease in oxygen availability, cells activate a specific gene expression pattern, under the control of the hypoxia inducible factors (HIFs), that mediates a set of stereotypical adaptations including an extensive metabolic reprogramming and the induction of mechanisms to increase oxygen delivery. Given the impact of these responses on the maintenance of tissue homeostasis, it stands to reason that the activation of the HIF pathway and downstream targets could contribute to the clinical progression of those diseases that course with hypoxia.

HIF is a heterodimer composed of an oxygen-regulated alpha subunit (HIFalpha) and a constitutively expressed beta subunit (Aryl Receptor Nuclear Translocator, ARNT,

*To whom correspondence should be addressed. Tel: +34 91 585 44 40; Fax: +34 91 585 44 01; Email: luis.peso@uam.es

also known as HIFbeta) (2,3) that partners with a number of basic helix-loop-helix transcription factors. There are three separate genes encoding for HIFalpha subunits, *HIF1A*, *EPAS1* and *HIF3A*. Although these genes differ in their expression pattern and biological activities, they are subjected to similar regulation by oxygen. Under normoxia, HIFalpha is efficiently hydroxylated at two proline residues (4,5) by a family of dioxygenases (EGL Nine homologs, EGLNs, also known as Prolyl hydroxylases, PHDs) that require oxygen as co-substrate (6,7). This post-translational modification is recognized by an E3-ubiquitin ligase complex, containing the von Hippel-Lindau protein (pVHL) (8), which targets HIFalpha for proteosomal degradation. Thus, under normal oxygen tension, HIFalpha half-life is extremely short and, consequently, normoxic protein levels are very low. In addition, another dioxygenase (Factor Inhibiting HIF, FIH) catalyses the oxygen-dependent hydroxylation of an asparagine residue, located in the C-terminal transactivation domain, preventing its interaction with the p300 coactivator and blunting HIFalpha transcriptional activity (9). Under hypoxia, all these hydroxylation reactions become compromised, due to the reduced availability of oxygen and the high Km of these enzymes for oxygen (10). As a consequence, HIFalpha becomes stabilized and free to interact with transcriptional co-activators such as p300. HIFalpha accumulation allows its interaction with ARNT subunit and binding of the heterodimer to the RCGTG motif, known as hypoxia response element (HRE), present in the regulatory regions of HIF target genes.

Genetic alteration of almost all the components of the HIF pathway have been described and linked to different pathological conditions (11,12). Specifically, loss of function mutations in the *VHL* tumor suppressor gene are associated with the von Hippel-Lindau disease, a hereditary cancer syndrome, (13) and a rare form of familial erythrocytosis known as Chuvash disease (14). Similarly, mutations altering *EGLN*s, in particular *EGLN1* (*PHD2*) and *EPAS1* have been linked to some familial cases of erythrocytosis (11,15–17) and to neuroendocrine tumors such as paragangliomas and pheochromocytomas (15,18). In addition to these links to human disease, polymorphisms in several of these genes, most notably in *EGLN1* and *EPAS1*, are responsible for the adaptation to the chronic hypoxia of high-altitude of some human populations (12,19,20). In spite of this wealth of information regarding the genetics of the components of the HIF signaling pathway and its contribution to human disease and evolution, little attention has been paid to the effect of variants in regulatory regions bound by HIF. Single nucleotide polymorphisms (SNPs) that disrupt the RCGTG motif prevent the transcriptional response to hypoxia as we demonstrated for the SNP rs17004038, mapping to a functional HRE in the promoter of the macrophage migration inhibitory factor gene (21). Moreover, the genetic variability of non-coding regions could have a strong impact on disease progression as demonstrated by Schodel *et al.* who described a polymorphism that modulates the binding of HIF to an enhancer that controls the transcription of the gene encoding for cyclin D1 in renal cell carcinoma (22).

Chronic obstructive pulmonary disease (COPD) and obstructive sleep apnea syndrome (OSAS) are two highly prevalent respiratory disorders (23,24) which are associated with a sustained or intermittent reduction in blood oxygen saturation and tissue hypoxia (25,26). HIF is induced in cellular and animal models of both COPD (27) and OSAS (28,29) and mediates the development of pulmonary (30–32) and arterial hypertension (33), which are two comorbidities associated to COPD and OSAS respectively that complicate their clinical management. In keeping with these studies in animal models, specific mutations in *EPAS1* have been associated to pulmonary hypertension (34,35) and a recent genome-wide analysis of DNA methylation and gene expression profiles of COPD lung samples identified EPAS1 as a key regulator of COPD (36). Altogether these evidences strongly suggest a role for the HIF pathway in the progression of COPD and OSAS.

Herein we sought to identify genetic variants affecting RCGTG motifs within HIF binding regions (HBR) in genomic samples from COPD and OSAS patients and determine their functional impact on the transcriptional response to hypoxia.

MATERIALS AND METHODS

Study subjects

Patients between 35 and 80 years of age admitted to the Respiratory Services of Hospital Universitario La Paz or Hospital Universitario de la Princesa (Madrid, Spain) with a diagnosis of moderate-very severe COPD or OSAS were considered for this study. The diagnosis of COPD was based on clinical history and spirometry criteria according to the Global initiative for chronic Obstructive Lung Disease (GOLD) guidelines. The inclusion criteria were a post-bronchodilator FEV1/FVC ratio of less than 0.7 and a post-bronchodilator FEV1 < 80% predicted. Patients suffering from bronchial asthma, diffuse interstitial lung disease, neuromuscular or chest wall diseases or lung neoplasia were excluded. A total of 96 COPD patients were recruited and classified according to the severity, clinical phenotype (37) and and/or presence of comorbidities. Regarding disease severity, patients were classified according to: degree of airflow limitation according to GOLD criteria; the annual number of exacerbations of the disease, defined as an acute worsening of the patient's baseline dyspnea, cough and/or sputum production and the multifactorial BODE (Body-mass index, airflow Obstruction, Dyspnea, and Exercise) index (38). As for the comorbidities, patients were classified based on previous diagnosis or by the following criteria: pulmonary hypertension, defined as an estimated pulmonary arterial systolic pressure >35 mm Hg; Arterial hypertension, as resting blood pressure is persistently at or above 140/90 mmHg in three consecutive determinations or 24-h ambulatory blood pressure above 135/85 mmHg; cardiovascular disease risk was defined by the co-occurrence of arterial hypertension, diabetes mellitus (current treatment with oral anti-diabetic drugs and/or insulin; a fasting glucose value above 126 mg/dl on at least two occasions; blood glucose level at 2 h after an oral glucose tolerance test is equal to or more than 200 mg/dl; or a glycated hemoglobin (HbA1c) level > 6.5%), dyslipidemia (total cholesterol ≥ 200 mg/dl, triglycerides ≥ 180 mg/dl, HDL-cholesterol ≤ 40 mg/dl or LDL-cholesterol ≥ 150 mg/dl), or metabolic syndrome (as defined by National Cholesterol Education

Program Expert Panel on Detection, Evaluation and Treatment of High Blood Cholesterol in Adults) (39); ischemic heart disease, was diagnosed following the European Society of Cardiology/American College of Cardiology guidelines (40); neoplasia, patients with a clinical diagnosis of any kind of neoplasm confirmed by histology. In addition, comorbidities were assessed globally according to the Copd cO-morbidity TEst (COTE index) (41).

The diagnosis of OSAS was based on evidence of excessive daytime sleepiness (Epworth Sleepiness Scale score > 10) and an Apnea–Hypopnea Index (AHI) > 5/h, determined by respiratory polygraphy or polysomnography, with at least 80% of the apneas being obstructive. The exclusion criteria were previous or current diagnosis of secondary hypertension, bronchial asthma, diffuse interstitial lung disease, neuromuscular or chest wall diseases, lung neoplasia, restrictive lung disease, bronchiectasis, alterations of thyroid function and morbid obesity (BMI > 40 Kg/m2). A total of 144 OSAS patients were recruited and classified according to the presence of comorbidities as described before for COPD and including the following additional classes: pulmonary embolism, diagnosed by angio-computed tomography; depression and anxiety, according to Beck depression inventory or State-Trait Anxiety Inventory; and heart failure, with echocardiographic confirmation of left ventricular systolic or diastolic dysfunction.

The study was approved by the La Paz Hospital Medical Ethics Committee (PI-795) and informed written consent was provided by all subjects.

High-throughput sequencing and analysis

DNA was isolated from whole blood using Gentra Puregene Blood Kit (Qiagen, 158467) and regions of interest were polymerase chain reaction (PCR)-amplified using a 48.48 Fluidigm Access Array (Fluidigm Corporation, South San Francisco, USA). The resulting bar-coded amplicons were sequenced in a MiSeq apparatus (Illumina) under a 2×250 pair-ended format (Genomics Units, Science Park, Madrid). Reads were quality filtered according to the standard Illumina pipeline, de-multiplexed and fastq files were generated.

Raw sequences generated by MiSeq (FASTQ format) were aligned to the February 2009 (GRCh37/hg19) assembly of the human genome with Bowtie2 (version 2.1.0) (42,43). Alignments were processed using Samtools (version 0.1.18) (44) and variants were called with bcftools filtering out those with mapping quality <30 or DP < 16. File manipulation and analysis were performed with custom scripts written in python (Python Software Foundation. Python Language Reference, version 3.4.3. Available at http://www.python.org) and R (R Core Team (2014). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.) languages. We used Python version 3.4.3 and R version 3.2.2. For the visualization of aligned reads (BAM files) we used the Integrative Genomics Viewer, Integrated Genomic Viewer (IGV) (45). For the treatment of genomic intervals we used the IRanges and Granges (46) libraries from Bioconductor version 3.2 (https://www.bioconductor.org/). The regions selected for

binding along with the supporting experimental evidence can be interactively accessed at the UCSC genome browser: http://genome.ucsc.edu/cgi-bin/hgTracks? hgS_doOtherUser=submit&hgS_otherUserName= Lab252&hgS_otherUserSessionName=Roche_et_al_2016.

Cell culture

Human Embryo Kidney 293T (HEK293T) and HeLa cells were grown using standard culture conditions in Dulbecos's Modified Eagle medium supplemented with 10% Fetal Calf Serum (FCS). For hypoxia treatments, cells were placed in a $1\% O_2$, $5\% CO_2$, $94\% N_2$ gas mixture in a Whitley hypoxystation (don Whitley Scientific, UK).

RNA extraction and quantitative RT-PCR

Total RNA was extracted and purified with the RNeasy Mini Kit (Qiagen) and treated with RNase free Dnase set (Qiagen, 79254). One microgram of total RNA of each sample was reverse-transcribed to cDNA (Transcriptor First Strand cDNA Synthesis kit, Roche) and 2 μl of a 1:20 dilution of this reaction was used as template for amplification reactions using Power SYBR green PCR Master Mix (Applied Biosystems, 4367659) or Taqman Universal Master Mix II (Applied Biosystems, 4440040), following the manufacturer's instructions and the primers/probes indicated in Supplementary Table SV. PCR amplifications were carried out in a StepOne Real-time PCR System (Applied Biosystems) and data were analyzed with StepOne software. Each sample was tested in duplicate, and expression levels calculated using $\Delta\Delta Ct$ method using β-actin as a reference.

Metabolic labeling with 4-thiouridine and purification of newly synthesized mRNA

We used the protocol described by Dolken *et al.*, and Schawnhausser *et al.* (47,48). Briefly, exponentially growing HeLa cells were exposed to 21 or 1% oxygen for 8 h and pulse-labeled with 4-thiouridine (400 μM, 4sU, Sigma, T4509) during the last two hours of treatment. After treatment, total RNA was isolated from cells using TRI reagent (Ambion, AM9738) and subjected to a biotinylation reaction to label the newly transcribed RNA containing the 4sU moiety. Next, RNA was extracted using Ultrapure TM Phenol:Chloroform:Isoamylalcohol (Invitrogen, 15593-031) and labeled RNA was isolated from the total RNA by affinity chromatography using streptavidin coated magnetic beads (μMacs Streptavidin Kit; Miltenyi, 130-074-101).

Gene knock-down

Cells were transfected with the indicated siRNAs at 100 nM concentration using Lipofectamine RNAiMax (Invitrogen), following the manufacturer's instructions and the knockdown efficiency was determined by real time quantitative PCR and western blot analyses. We used commercial siRNAs, *HIF1A* siRNA (Santa Cruz, sc-44225), *EPAS1* siRNA (Santa Cruz sc-35316) and siRNA-B (Santa Cruz sc-44230) as a negative control.

Plasmid constructs

Human genomic DNA, extracted from peripheral blood from individuals heterozygous for the variants of interest, was used as template for PCR amplification using the primers described in Supplementary Table SIII. Target regions were delimited based on the information about Dnase sensitivity, clustering of transcription factor binding sites and histone modifications associated to transcriptionally active regions (H3K27Ac, H3K4Me1 and H3K4Me3) generated by the ENCODE project (49). Reporter constructs were generated by cloning the PCR products into the pGL4.20 or pGL4.23[luc2/minP] reporter vectors (Promega). Both, reference and alternate, alleles were cloned and the identity of all constructs was verified by Sanger sequencing. The constructs containing the mutated regulatory region of *EGFR* and *PGK1* were generated by site-directed mutagenesis, employing PCR QuikChange IIXL Site-direct mutagenesis kit (Agilent) and the primers described in Supplementary Table SV.

Reporter assays

HeLa cells were plated in six-well plates 24 h prior transfection (200 000 cells/well, six-well plates). Each well was transfected with a DNA mixture containing 1 μg of the indicated reporter plasmid, 0.033 μg of a plasmid encoding the Renilla firefly luciferase under the control of a CMV promoter and 7.967 μg of pCDNA3. Twenty-four hours after transfection, cells were split into 24-well plates and then transferred to hypoxic conditions (1% oxygen) or left under normoxic conditions for 16 h. Subsequently, firefly and renilla luciferase activities were determined using a *Dual Luciferase Reporter 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 triplicate.

Genome editing

HEK 293T cells were edited by CRISPR-Cas9 to disrupt the HRE elements within the EGFR and EGLN3 loci. Single-guide RNAs (sgRNAs) were designed using CRISPR Design Tool from Feng Zhang's Lab. (http://tools.genome-engineering.org) and four top scoring sequences per loci were cloned into PX459 (Addgene, Cambridge, MA, USA). Efficiency of sgRNAs was checked by SURVEYOR assay (IDT, Coralville, Iowa, USA), and the best sequence used in subsequent experiments (Supplementary Table SVI). For editing, HEK 293T cells were transfected with 2.5 μg of PX459 and 0.6 μg of single-stranded DNA oligonucleotides as repair template containing the intended nucleotide substitutions (Supplementary Table SVI) using Lipofectamine® 2000 (Thermo Fisher Scientific). Bulk cultures were plated clonally at limiting dilution in 96-well plates. Clones were expanded and genotyped by Sanger sequencing.

RESULTS

Selection and characterization of genome-wide HIF binding regions

In order to identify genetic variants that could affect the transcriptional response to hypoxia, we limited our search to HIF binding regions (HBR) across the genome. To define a comprehensive list of these regions, we borrowed data from studies describing the genome-wide pattern of binding of HIF1A and EPAS1, determined by means of chromatin immunoprecipitation followed by massive parallel sequencing (ChIP-Seq) or hybridization to high-density oligonucleotide tiling microarrays (ChIP-Chip) (22,50–53). The comparison of these data sets revealed a very limited overlap between the HIF binding regions identified in these studies, with <20% of regions being common to any pair of assays (Figure 1A). This result is probably a consequence of a combination of biological factors, including variability in cell-type and HIF-isoform specific targets, and also of technical differences across these studies. Regardless the explanation, the reduced overlap implied that we could not derive a comprehensive list of HBR from the intersection of these data sets. On the other hand, the number of nonredundant regions increased linearly with the number of studies up to circa 1700 binding regions (Figure 1B), so that cost-efficiency considerations precluded us from analyzing the complete set of all unique regions. In addition, some of the sites identified in a single dataset could be false positives as spurious binding sites are not expected to be reproduced across studies. Thus, to reduce the chances of including false positives, we took into account the effect of hypoxia on gene expression as an additional source of information to guide us discriminate *bona fide* HBR. To this end, we ascribed each of the regions from the ChIP studies to the nearest flanking gene and their response to hypoxia was assessed based on a published meta-analysis of expression profiles of cells exposed to hypoxia (21). Conversely, to preserve sensitivity, we complemented the set of unbiased studies with a list of experimentally validated HREs derived from a manual inspection of bibliography (see Supplementary Table SI). Finally, we derived the list of regions to be sequenced by the application of a set of rules to the compiled collection of HIF-binding datasets (Figure 1C and Supplementary Table SI). Specifically, to increase specificity we only selected those regions that were bound by HIF in at least three independent datasets. On the other hand, aiming to generate a list as exhaustive as possible, we also selected those regions that, while being bound by HIF only in one or two of the datasets, were associated to a gene significantly induced by hypoxia according to the gene expression profiles meta-analysis. Finally, we also included a few regions derived from individual gene studies that, probably due to their cell-type specific pattern of expression (e.g. *EPO*), were neither found in ChIP datasets nor labeled as hypoxia-responsive in the meta-analysis (see Supplementary Table SI). The application of this strategy (Figure 1C) resulted in a total of 238 HBR regions (Supplementary Table SI). The majority of these regions were bound by HIF in two or more independent unbiased (i.e. those derived from ChIP studies) datasets (Figure 1D). Fi-
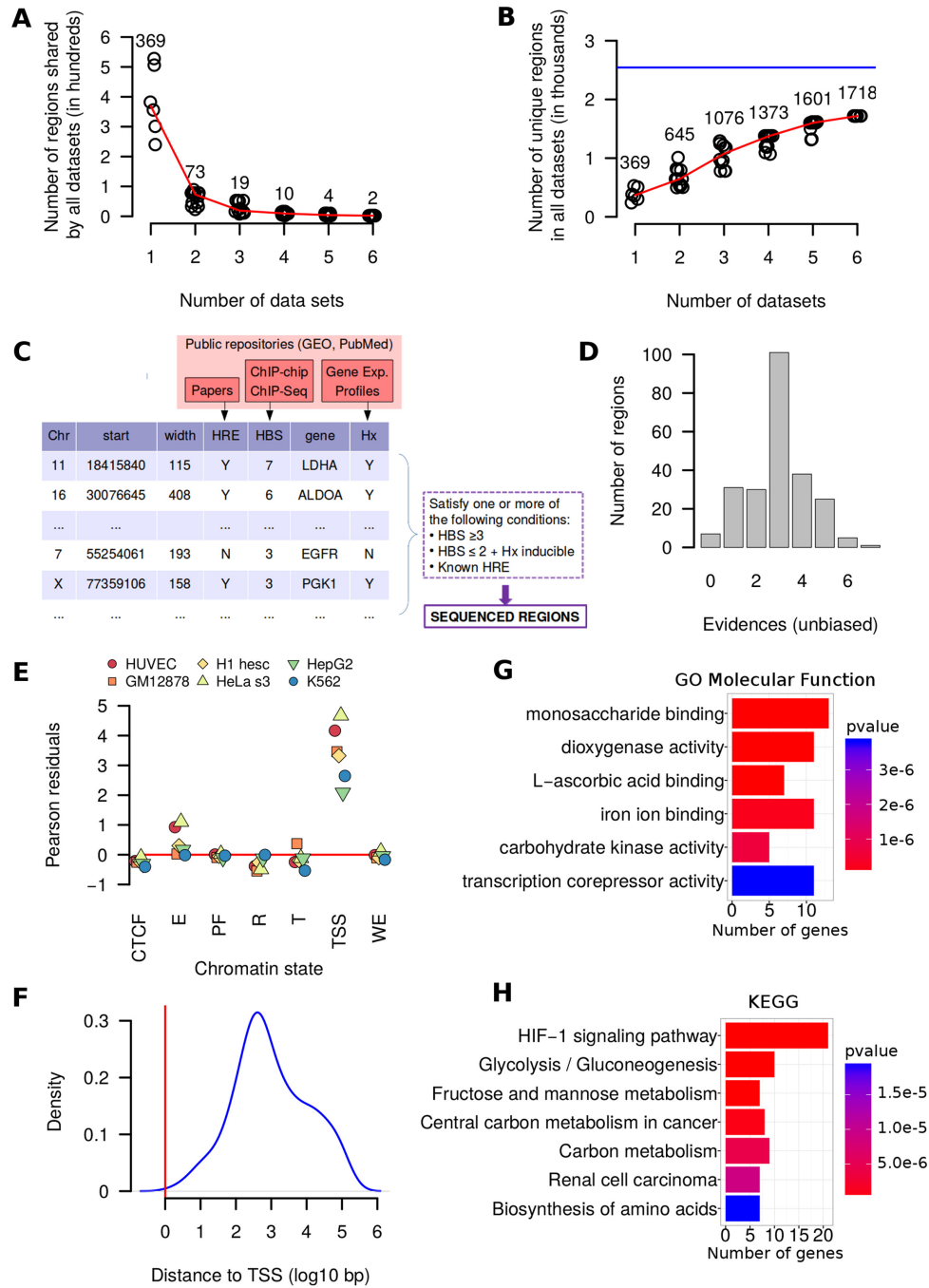
Figure 1. Identification and characterization of HIF binding regions (HBR). (**A** and **B**) HIF binding sites from references (22,50–53) were compared to determine their degree of overlap (A) and redundancy (B). The indicated number of datasets (x axis) were selected at random from a total of six tables of HIF binding sites and the number of regions overlapping across all selected tables (A) or number of independent non-overlapping regions (B) were recorded. The graph represent the results from 10 independent samplings, the number above the points is the median value of the 10 samplings and the line joins these median values. The the top horizontal line in (B) indicates the value of the sum of the number of HIF binding regions in all the datasets. (**C**) Schematic representation of data sources and rules followed to select the set of HBR. (**D**) The histogram represents the distribution of the number of HIF binding sites identified in ChIP experiments overlapping each of the HBR selected for sequencing. We considered the HIF binding sites described in the references indicated in (A) plus those in reference (62). (**E**) The distribution of the HBR across the functional regions of the genome was studied in the indicated cell lines and compared to that expected by chance using a chi-squared goodness of fit test ($P < 0.001$ for all cell lines). The graph represents the Pearson residuals resulting from the test in each cell line. The functional regions (chromatin states as defined by the ENCODE consortium, (49)) are: 'TSS', predicted promoter region including TSS; 'PF', Predicted promoter flanking region; 'E', Predicted enhancer; 'WE', Predicted weak enhancer or open chromatin cis regulatory element; 'CTCF', CTCF enriched element; 'T', Predicted transcribed region; 'R', Predicted Repressed or Low Activity region. (**F**) Distribution of distances of the HBRs to the nearest transcription start site. The vertical line marks the TSS position. (**G** and **H**) Enrichment of Gene Ontology (GO) terms (G) and KEGG pathways (H) was investigated for the genes nearest to the HBRs using the R package 'clusterProfiler' (71). The figure shows the categories found enriched with an FDR-adjusted *P*-value < 0.001. Bar length represent the number of genes bound by HIF in each category and the shade of color the raw *P*-value for the enrichment.

nally, in order to meet the restrictions of the library and sequencing platforms, these HBR were trimmed to 250-bp long regions. Trimming was made based on Dnase hypersensitive regions determined by the ENCODE project (49) and, when necessary HBRs were split into two 250-bp long regions.

Next, we sought to characterize the selected HBRs. First, we investigated the distribution of these regions across different functional elements defined by the genome-wide segmentation in six different cell lines (GM12878, K562, H1-hESC, HeLa-S3, HepG2 and HUVEC) as defined by the ENCODE project (49). As shown in Figure 1E, the selected HBRs clustered in promoter regions (TSS) and, to a lesser extent, enhancer regions (E), while being excluded from repressed (R) and CTCF-enriched regions (CTCF); a distribution that was significantly different to that expected by chance (Chi-square$_6$ > 10^6, P < 0.0001). In agreement with this preference for promoter-like regions, the distribution of distances of the HBRs to the nearest transcription start site, peaked at circa 500 bp (Figure 1F). These results are congruent with the distribution of HIF binding sites described for individual cell lines (51–53). Lastly, we studied the ontology terms of the genes adjacent to the selected HBRs and found that they were significantly enriched in terms distinctive of biological process regulated by hypoxia such as carbohydrate metabolism, oxidation-reduction and iron metabolism (Figure 1G and H). Taken together, these evidences support the biological relevance of the selected set of HBR.

Identification of genomic variants affecting RCGTG motifs within HBRs

To identify genetic variants that could affect HIF-mediated transcription and contribute to chronic respiratory disease progression, we obtained genomic DNA samples from of 96 COPD and 144 OSAS patients and sequenced the 238 HBRs defined above in these samples by parallel massive sequencing. The resulting reads were aligned to the February 2009 assembly of the human genome (hg19, GRCh37) and genetic variants called using standard software. This analysis revealed a total of 458 and 569 variant loci in COPD and OSAS patients respectively as compared to the reference genome. Most of the observed changes (83 and 96% respectively) comprised single nucleotide variants (SNV), but we also observed small insertions-deletions (INDELs). Visual inspection of the aligned reads using the IGV revealed that, while SNV calls were robust, the annotation of INDELs was not always reliable. Accordingly, we disregarded INDELs and for the remaining of our analysis we focused on the 379 and 548 SNVs found in COPD and OSAS patients respectively (Figure 2A). Taken together, all the variants found in the 240 sequenced samples, added up to a total of 715 unique SNVs (Figure 2A). However, only 176 of these (25% of total) were frequent variants, defined here as those whose minor allele was present in more than 2% of the sequenced samples (Figure 2A, 'Frequent variants'). As expected, the vast majority of these frequent variants corresponded to SNPs previously described and present in the built 142 of the dbSNP database (Figure 2B). To increase the chances of identifying variants with functional impact

on the transcriptional response to hypoxia, we further narrowed our analysis to those SNV that altered the sequence of ACGTG or GCGTG motifs located within the HBRs. As shown in Figure 2C, of the unique 715 SNV identified across the 238 HBR in the 240 sequenced individuals, only 57 affected the sequence of RCGTG motifs and, of these, only 13 were present in more than 2% of all the samples (frequent variants). All these 13 SNV correspond to variants present in dbSNP142 (Figure 2D). The complete list of variants that result in the alteration or disruption of RCGTG motifs within HBRs is shown in Supplementary Table SII.

Next, we studied whether the presence of these variants correlated with the clinical features observed in patients. In the case of COPD patients we studied the association between the 12 frequent SNVs that disrupted the HREs (Supplementary Table SII) and their clinical phenotype (non-exacerbator, mixed COPD-asthma, exacerbator with emphysema and exacerbator with chronic bronchitis), comorbidities (cancer, cardiovascular risk, arterial hypertension, ischemic heart disease, pulmonary artery hypertension/cor pulmonale) and the stage of the disease according to different grading systems (severity of airflow limitation and BODE and COTE indices). As shown in Figure 3A, we found some weak associations with marginally significant statistics (*P*-values slightly below 0.05). However, upon correction for multiple-testing, none of the associations reached significance (Supplementary Figure S1A). Similarly, the analysis of the association between the 14 frequent SNVs found in OSAS patients that disrupt the RCGTG motif and the incidence of comorbidities (cancer, systemic hypertension, ischemic heart disease, heart failure, pulmonary artery hypertension, pulmonary embolism, diabetes mellitus, dyslipidemia, metabolic syndrome and depression-anxiety) revealed some weak correlations (Figure 3B) that were not significant (Supplementary Figure S1B) after correction.

In summary, our analysis identified a set of 14 frequent SNVs that disrupt the consensus HIF binding motif in COPD and OSAS samples (Supplementary Table SII). Although none of them showed a significant association with comorbidities or clinical phenotypes, we cannot rule out that a larger sample size, in particular for low-frequency phenotypes, could yield sufficient evidence to support an association. On the other hand, these variants map to RCGTG motifs within HBR and thus, at least some of them, are likely to result in an altered transcriptional response that might contribute to phenotypes not included in our analysis.

Characterization of the functional impact of genomic variants on HIF-mediated transcription

In view of these results, we next studied the effect of these SNVs on the transcriptional response to hypoxia. Based on the information on chromatin accessibility, histone modifications and clusters of transcription factor binding sites generated by the ENCODE project (49), we defined regulatory regions that contained each one of the 14 common SNVs that disrupt RCGTG motifs. Since variants chr10:3089315G>A (rs7078831) and chr10:3089424G>A (rs6602019) were very close to each other, we defined
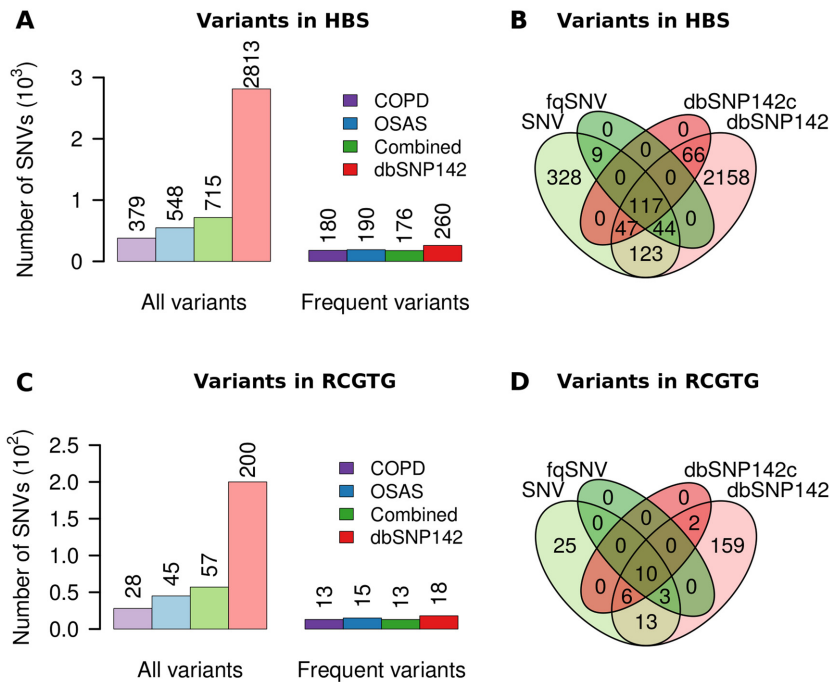
Figure 2. Genomic variants affecting HIF binding regions (HBRs). The 238 HBRs described in Figure 1 and Supplementary Table SI were sequenced in 96 COPD and 144 OSAS patients. The resulting reads were aligned to the human reference genome (GRCh37/hg19 assembly) to identify single nucleotide variants (SNVs). The figure represent the total number of SNV ('All variants') and the number of SNVs with minor alleles found in more than 2% of the samples ('Frequent variants') from COPD, OSAS and all patients combined that map to the HBRs (**A**) or to the RCGTG motifs within HBRs (**C**). The number of SNPs from the dbSNP142 database mapping to the HBRs is also represented. In this case the number in 'Frequent variants' correspond to the 'common SNP' subset of the dbSNP142 database containing variants with MAF > 1%. The overlap between the variants found in our study (all, 'SNV'; frequent, 'fqSNV') and the SNPs recorded in dbSNP142 (all SNP in database, 'dbSNP142'; common SNPs, 'dbSNP142c') is shown for all variants in HBRs (**B**) or those mapping to RCGTG motifs (**D**).
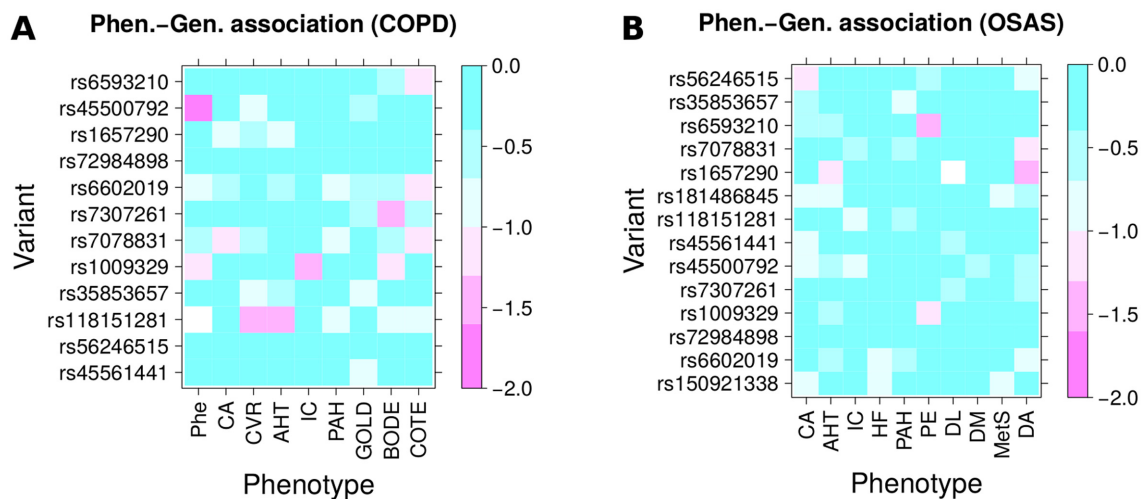

Figure 3. Association between genomic variants and clinical features of disease progression. We constructed a contingency table for each combination of frequent SNV and clinical features and applied a Fisher's exact test to each table. Each cell in the matrix represents the resulting P-value ($-$log10 of values) for the comparison of the genotype on the y-axis and the phenotype on the x-axis. (**A**) Analysis of the 12 frequent SNV disrupting RCGTG motifs found in COPD patients against their clinical phenotype (Phe), disease stage (according to the GOLD, BODE and COTE classifications) and comorbidities. For the analysis BODE and COTE categories (0–10) were binned into groups (Q1: BODE 0–2, Q2: BODE 3–4, Q3: BODE 5–6, Q4: BODE 7–10; COTE < 4 and COTE ≥ 4). (**B**) Analysis of the 14 frequent SNV disrupting RCGTG motifs found in OSAS patients against their comorbidities. CA: Cancer at any location; CVR: Cardiovascular risk; AHT: Systemic hypertension; IC: Ischemic heart disease; PAH: Pulmonary artery hypertension; HF: Heart failure; PE: Pulmonary embolism; DM: Diabetes mellitus; DL: Dyslipidemia; MetS: Metabolic syndrome; DA: Depression-anxiety.

a single genomic interval containing both variable positions. Similarly, the pair chr11:71010448C>G (rs35853657) and chr11:71010489G>C (rs181486845) also belong to the same enhancer region and were included in a single genomic interval for their analysis. Thus, we identified a total of 12 independent regulatory regions containing polymorphic RCGTG motifs (Supplementary Table SIII and Figure S3). The gene loci nearest to each of these regions were *ATP9A*, *BCKDHA*, *CTDP1*, *DNAJB6*, *EGFR*, *INSIG1*, *PFKP*, *PGK1*, *PPME1*, *RNMT*, *SHANK2* and *SLC2A3* (Supplementary Table SIII). Next, we selected DNA samples from patients heterozygous for each of these variants and cloned the regulatory segments from both alleles into reporter vectors. We were able to clone 10 out of the 12 regions and found that most of them (8 out of 10), were significantly regulated (single sample student's *t*-test, $P <$ 0.05) by hypoxia to some extent (Figure 4). HIF binding has been associated to upregulation of gene expression rather than repression (21,51,53); accordingly, hypoxia led to an induction of transcription in all the cases except for the region containing variant rs118151281 (ATP9A) whose activity was downregulated by hypoxia (Figure 4). Comparison of the luciferase activity in samples transfected with constructs containing the reference and alternate alleles revealed that two of the variants, rs1009329 (DNAJB6) and rs6593210 (EGFR) completely abrogated the response to hypoxia (Figure 4), whereas an additional variant, rs150921338 (PGK1) had a partial, yet significant (paired student's *t*-test, $P < 0.05$), effect on the induction of the reporter gene by hypoxia (Figure 4).

To gain further insight into the effect these variants, we selected two of them rs150921338 and rs6593210, located within the *PGK1* and *EGFR* loci respectively, for further analysis. *PGK1*, encoding for the isoform 1 of the glycolytic enzyme phosphoglycerate kinase 1, was one of the first HIF target genes to be described (54). Reassuringly, this variant results in a change T>A that destroys an ACGTG motif in the proximal promoter of the *PGK1* that was implicated in the induction of this gene by hypoxia (54). However, this variant affects one of three RCGTG motifs that have been reported to be functional in the *PGK1* promoter region (54) (Figure 5A), which could explain its partial effect on the induction of this gene (Figure 4). In agreement with this possibility, mutation of each individual HRE in the *PGK1* promoter revealed that the HRE affected by rs150921338 ('HRE2' in Figure 5B) has only a partial contribution to the induction of the construct by hypoxia and a second RCGTG in the promoter ('HRE1' in Figure 5B) had a larger effect on the transcriptional response (Figure 5B).

On the other hand, variant cchr7:55254186G>A (rs6593210) lies in an intron of the Epidermal Growth Factor Receptor gene, in a region containing functional features typical of enhancer regions (Figure 6A). Although the regulation of the expression of this receptor by oxygen levels has been previously described, it has been regarded as a post-translational effect of hypoxia on the translation rate of the mRNA encoding for it (55). For these reasons, we first decided to investigate the effect of hypoxia on the transcription of this gene. As shown in Figure 6B, hypoxia led to an increase of *EGFR* mRNA in the total mRNA fraction of similar

magnitude to the one observed for *PGK1*. In contrast, we did find any effect on *RUVBL2* mRNA, a transcript whose expression level does not respond to changes in oxygen tension (56). An increase of total mRNA could be due to either, augmented transcription of the gene encoding for it or reduced decay rate. To differentiate between these possibilities, we pulse-labeled cells with a 4-thiouridine (4sU) to label nascent RNA and determined the levels of *EGFR* by quantitative PCR in the 4sU-enriched mRNA fraction (57). This experiment revealed that hypoxia led to an increase of newly transcribed mRNA, supporting a transcriptional effect of hypoxia on *EGFR* expression. As expected, hypoxia also induced the transcription of *PGK1* while it had no effect on *RUVBL2* (Figure 6B right panel). In addition, we found that the increased transcription of the EGFR gene was mediated by HIF as the interference of *HIF1A* and, in particular, *EPAS1*, prevented the accumulation of *EGFR* mRNA during hypoxia (Figure 6C). In agreement with this result, *EGFR* mRNA was significantly increased in clear cell renal cell carcinomas (ccRCC), a tumor type where VHL is frequently lost and, as a consequence, contain a constitutively active HIF pathway (Supplementary Figure S2). Altogether these results suggest that hypoxia induces the transcriptional upregulation of the *EGFR* gene through the direct binding of HIF to an intronic enhancer region (Figure 6A) and that variant rs6593210 prevents this regulation. To further test this hypothesis, we edited the genome of HEK293T, by means of the CRISPR-cas9 technology (58), to introduce a nucleotide change mimicking variant chr7:55254186G>A (rs6593210) and studied the effect of this modification on the transcriptional response to hypoxia. As a control for this experiment, we also introduced a disrupting mutation in the HRE motif that was shown to be responsible for the hypoxic induction of *EGLN3* and which is located within an intron more than 20 kb downstream of the transcription start site of this gene (59). Following transfection with the constructs targeting each loci, single cell clones were isolated, expanded and genotyped by Sanger sequencing. We screened 148 clones derived from the pool of cells engineered to contain the variant chr7:55254186G>A (rs6593210) in the *EGFR* locus (EGFR cell clones) and 90 clones from the transfection targeting the *EGLN3* HRE (EGLN3 cell clones). HEK293T is a hypotriploid cell line that contains three copies of the *EGFR* and *EGLN3* loci (60). Thus after discarding clones with ambiguous genotypes, we selected four EGFR cell clones in which at least one of the HRE copies was disrupted by either the introduction of the intended variant (genotype 'WT/SNP') or a small insertion-deletion (genotype 'WT/INDEL') created by non-homologous repair (Supplementary Table SIV). We also found two additional EGFR cell clones with all three HRE copies edited (genotypes 'SNP/SNP' and 'SNP/INDEL'). In the case of *EGLN3* locus we identified three cell clones in which at least one of the HRE copies was disrupted by either the intended mutation ('WT/MUT') or an INDEL ('WT/INDEL'), but none in which all alleles were simultaneously edited. In addition, since we expected inter-clonal variability in the response to hypoxia, we also isolated a total of nine additional clones that showed no alteration of the targeted loci to be
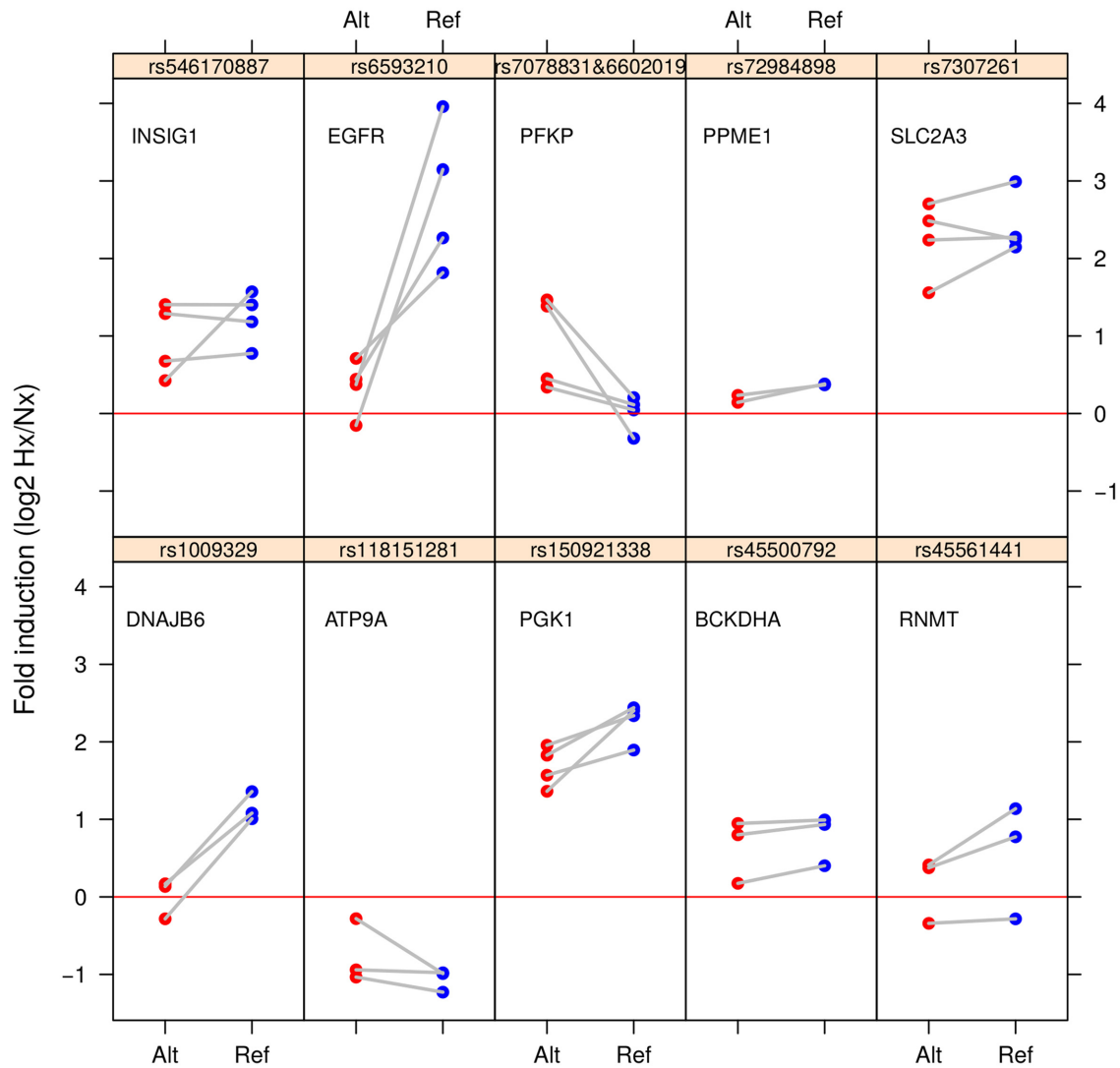
Figure 4. Effect of SNVs on the transcriptional response to hypoxia. HEK293T cells were transfected with reporter constructs containing the indicated variants or their corresponding reference alleles. After transfection, cells were exposed to normoxia or hypoxia for 16 h and processed to determine luciferase and renilla activities. Each panel represents the results for a regulatory region and the gene nearest to the cloned region is shown as a text inset. The graphs represent renilla-corrected luciferase activity in hypoxic samples expressed as log2-fold over the activity obtained in normoxic conditions. Each dot represent the results obtained in a single experiment and segments join the values corresponding to the construct containing the reference ('Ref') and alternate ('Alt') variants in the same experiment. The horizontal line marks the value of zero (no induction). The transcription mediated by regulatory regions containing the reference allele of the SNVs rs546170887 (*INSIG1*), rs6593210 (*EGFR*), rs72984898 (*PPME1*), rs7307261 (*SLC2A3*), rs1009329 (*DNAJB6*), rs118151281 (*ATP9A*), rs150921338 (*PGK1*) and rs45500792 (*BCKDHA*) were significantly modulated by hypoxia (single sample *t*-test, $P < 0.05$). The response to hypoxia of the reference and alternate variants was significantly different (paired *t*-test, $P < 0.05$) in the case of rs6593210 (*EGFR*), rs1009329 (*DNAJB6*) and rs150921338 (*PGK1*).

used as reference controls (Supplementary Table SIV). We then exposed wild-type (containing reference alleles) and edited cell clones to normoxia or hypoxia and determined *EGFR* and *EGLN3* mRNA levels by quantitative PCR to evaluate the impact of CRISPR-mediated alterations on the transcription of these genes.

To validate this approach, we first analyzed the consequences of editing the HRE on the induction of *EGLN3* mRNA by hypoxia (Figure 6D, left graph) and found that, in spite of inter-clonal variation, all wild-type clones showed a robust induction of *EGLN3* mRNA in response to hypoxia (mean induction of 4.2 times over normoxia). Importantly, all three EGLN3 cell clones with defective HRE alle-

les showed a reduced *EGLN3* response to hypoxia and the mean induction of *EGLN3* in this group was significantly different to that of the wild-type clones (Student's *t*-test t(4.7) = 5.88, $P < 0.01$). Moreover, as expected for clones containing some remaining functional copy of the HRE, the fold induction of *EGLN3* in the heterozygous cells was circa 0.5 times that observed in the wild-type clones (Figure 6D, left graph), suggesting that the targeted HRE is the only RCGTG motif that significantly contributes to the regulation of this gene by hypoxia.

Next, we investigated the effect of introducing variant chr7:55254186G>A (rs6593210) on the induction of *EGFR* mRNA by hypoxia. The wild-type EGFR cell
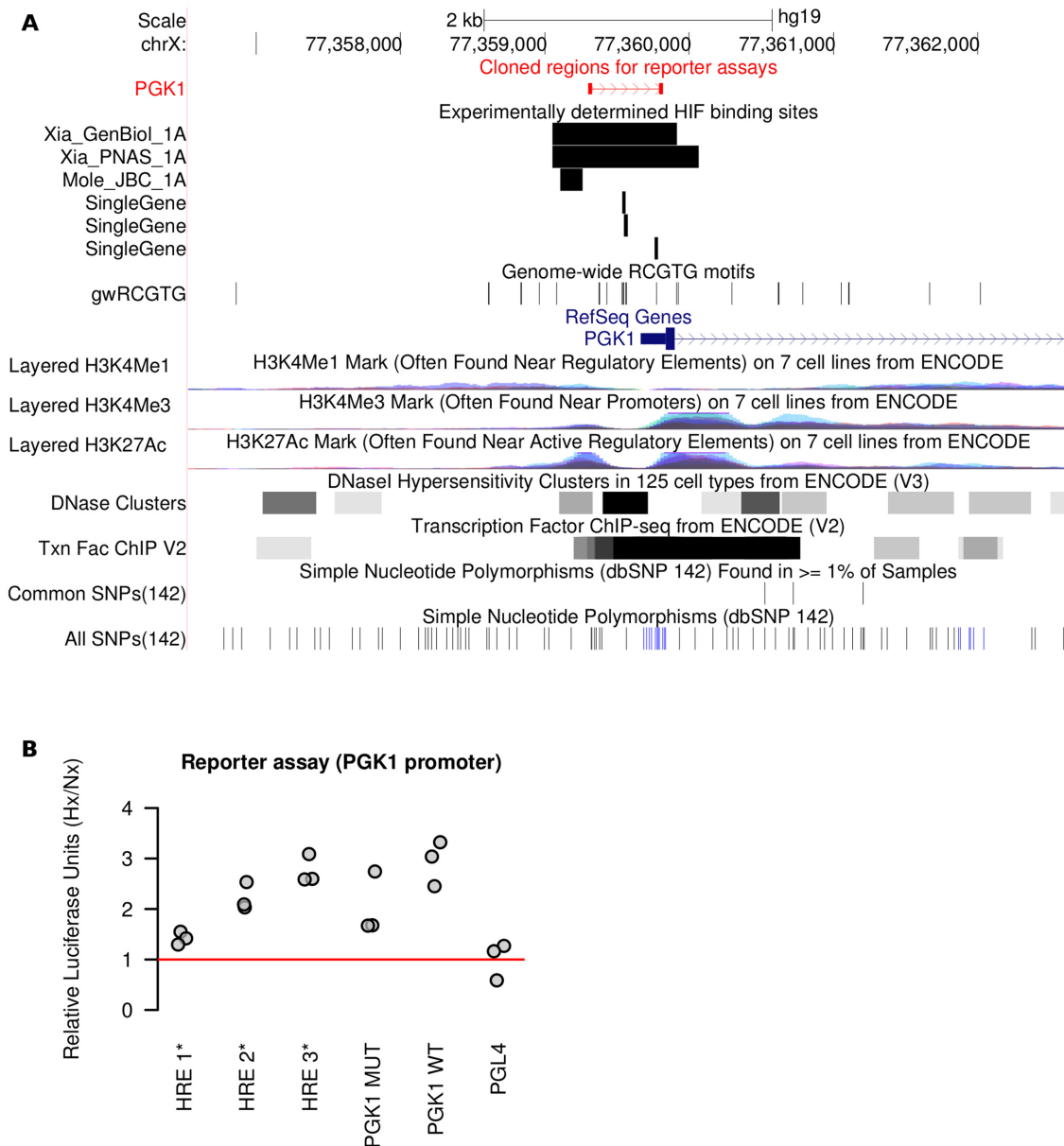
Figure 5. Transcriptional regulation of the *PGK1* gene. (**A**) Diagram depicting the promoter region of the *PGK1* locus and showing the RefSeq genes along with accessible chromatin regions ('DNase clusters' track), clusters of transcription factor binding sites ('Txn Fac ChIP' track), histone marks associated to regulatory elements ('Layered H3K4Me1' track), histone marks associated to promoters ('Layered H3K4Me3' track) and active regulatory elements ('Layered H3K27Ac Track'). The different histograms in the histone tracks correspond to the signal obtained in different cell lines (see UCSC for details). The figure was generated by the UCSC genome browser (72) upon loading custom tracks to indicate the location of HIF binding regions from the references indicated in Figure 1 ('Xia_GenBiol', 'Xia_PNAS', 'Mole_JBC') and single gene studies ('SingleGene'), the genome wide RCGTG motifs ('RCGTG' track) and the region cloned to assay its transcriptional activity ('PGK1' track on top of the diagram). (**B**) The effect of hypoxia on the transcriptional activity of the *PGK1* promoter region was assessed upon transfection of HeLa cells with constructs in which the luciferase was under the control of the wild-type sequence ('PGK1 WT'), a sequence containing variant chrX:77359566T>A (rs150921338) ('PGK1 MUT') or a mutant sequence with truncated HREs ('HRE1*','HRE2*', 'HRE3*', numbered from the ATG; HRE2 corresponds to the polymorphic RCGTG motif). The effect of hypoxia on the transcription of the empty reporter plasmid is also shown ('PGL4'). The graph represents the normalized luciferase activity in hypoxic samples expressed as fold over the activity obtained in normoxic conditions. Each symbol represent the results obtained in an independent experiment. The differences between groups were statistically significant (ANOVA $F_{3,11} = 55.1$, $P < 0.001$) and the mean induction of the WT group was significantly different to that of SNP and MUT (adjusted $P < 0.001$) in a *posteriori* Tukey test.
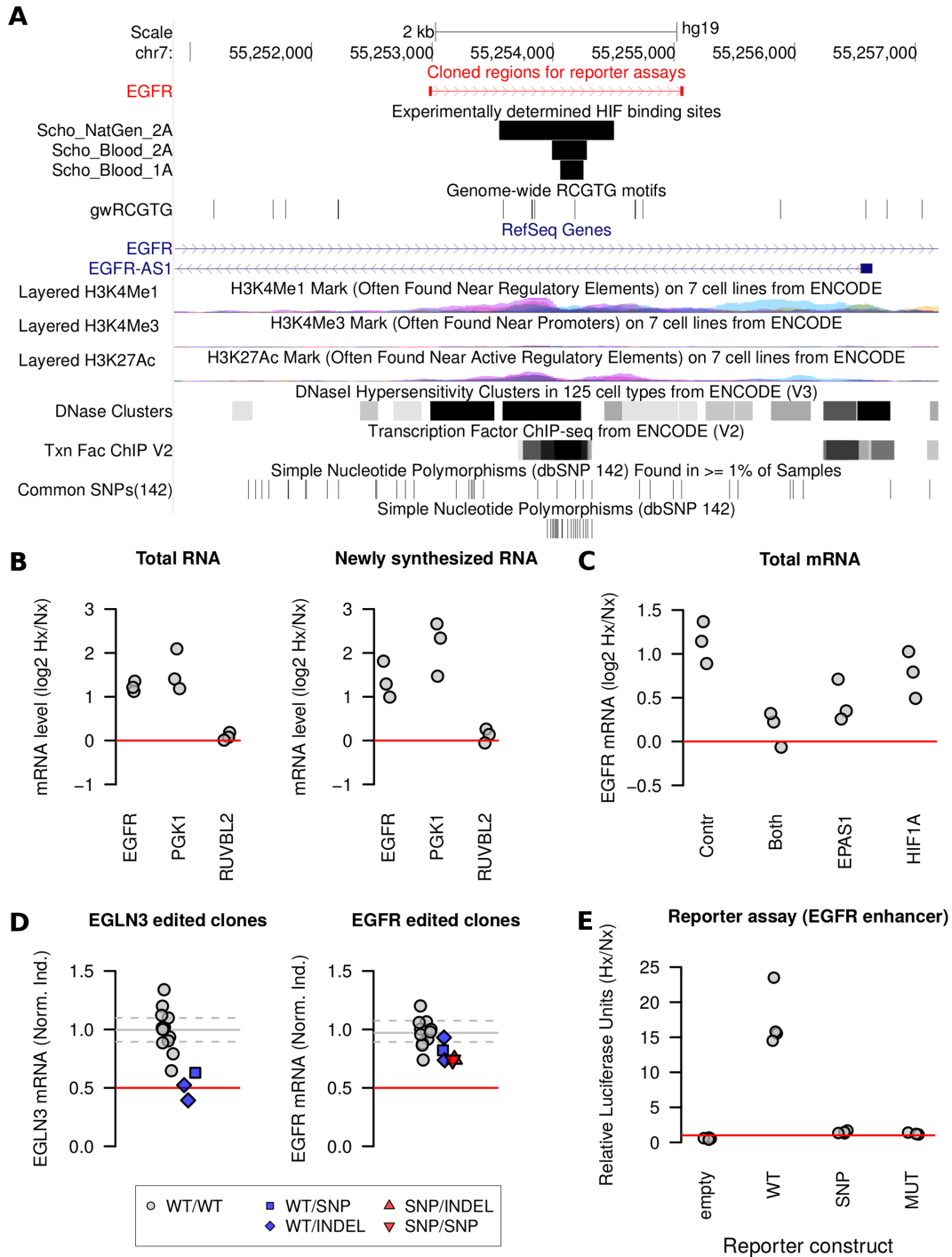
Figure 6. Transcriptional regulation of the EGFR gene. (**A**) Diagram depicting intron 22 of the EGFR locus and the region cloned to assay its transcriptional activity ('EGFR' track on top of the figure). Track names and labels are as indicated in Figure 5A. The binding regions of EPAS1 ('Scho_NatGen' and 'Scho_2') and HIF1A ('Scho_1') to this locus are shown as custom tracks. (**B**) HeLa cells were exposed to normoxia or hypoxia (1% oxygen) for 16 h and pulse labeled with 4sU during the last 2 h of treatment. Total RNA was extracted ('Total RNA') and an aliquot was subjected to affinity chromatography to purify 4sU-labeled ('Newly synthesized') fraction. Then the expression levels of *EGFR*, *PGK1* and *RUVBL2* were determined by quantitative PCR in each RNA fraction. The graph shows the ratio of hypoxic over normoxic RNA levels in log2-units. Symbols represent determinations from three independent experiments. The induction of *EGFR* and *PGK1* by hypoxia was statistically significant in both total (single sample t-test: $t_2 = 18$, $P = 0.003$ and $t_2 = 5.7$, $P = 0.029$ respectively) and newly synthesized (single sample t-test: $t_2 = 5.7$, $P = 0.029$ and $t_2 = 6$, $P = 0.026$ respectively) fractions. (**C**) HeLa cells were treated with the indicated siRNAs and then grown at normoxic (21% oxygen, Nx) or hypoxic (1% oxygen, Hyp) conditions for 12 h. The graph represent the ratio of normalized levels of *EGFR* RNA in hypoxic over normoxic samples in log2-units. Symbols represent determinations obtained in independent experiments. The differences between groups were statistically significant (ANOVA $F_{3,8} = 9.39$, $P < 0.001$) and the mean of groups 'EPAS1' and 'Both' were the statistically significant to that of the control group ($P < 0.05$ and $P < 0.01$, respectively) in a *posteriori* Tukey test. (**D**) HEK293T cells were edited to disrupt the HRE elements within the *EGFR* and the *EGLN3* loci and the genotype of RCGTG motif in each clone was determined by Sanger sequenc-

clones showed a modest, yet consistent, induction of *EGFR* mRNA by hypoxia (mean value of 2.5 times over normoxia). Importantly, the induction of *EGFR*, in all but one of the edited cell clones, was reduced and their response to hypoxia was outside the 95% confidence interval of the mean induction observed for the wild-type clones (Figure 6D, right graph). In fact, the mean fold induction of EGFR in clones with all HREs disrupted was significantly lower than that observed for clones with functional alleles (ANOVA $F_{2,14} = 4.987$, $P < 0.05$, pos-hoc Tukey test $P < 0.05$). These results indicate that the HRE affected by variant chr7:55254186G>A (rs6593210) significantly contributes to the transcriptional upregulation of *EGFR*. Nevertheless, although attenuated, the response of edited clones is still higher than expected (above 0.5 times of the wild-type response). Thus, it is likely that other HRE collaborate with the polymorphic HRE affected by rs6593210 to regulate *EGFR* in response to hypoxia. These ancillary unidentified elements must be located in a regulatory region distinct from the one cloned in our reporter constructs, because the response to hypoxia observed for the wild-type enhancer is completely abolished by the introduction of the single nucleotide change G to A mimicking the variant chr7:55254186G>A (rs6593210) (Figure 6E).

Altogether these results indicate that *EGFR* gene contains an enhancer in its 22nd intron that, probably in combination with other unidentified regulatory modules, contributes to the transcriptional induction of *EGFR* gene in response to hypoxia. Furthermore, the transcriptional activation of this enhancer by hypoxia is mediated by a single HRE whose activity is completely abrogated by the variant chr7:55254186G>A (rs6593210).

Frequent polymorphisms are excluded from functional HREs

The analysis presented above resulted in the identification of three variants rs1009329 (DNAJB6), rs150921338 (PGK1) and rs6593210 (EGFR) that affect the response to hypoxia. However, only two of them (rs1009329, rs6593210) completely prevented the reporter activity. On the other hand, we also found four variants (rs546170887, rs45500792, rs72984898 and rs7307261) that did not have a significant effect on hypoxia-driven transcription, implying that additional non-variant RCGTG motifs within the cloned regulatory regions mediate the induction of tran-

scription. Altogether, these results suggest that most polymorphic RCGTG motifs are either non-functional or redundant in the induction of transcription in response to hypoxia, hinting that frequent polymorphisms might be excluded from functional HREs. To test this possibility we investigated whether the proportion of polymorphic RCGTG motifs varied between motifs mapping inside and outside HBRs. As shown in Figure 7A, the proportion of RGCGT motifs affected by SNVs was significantly lower in regions bound by HIF (chi-square = 16.178, P-value = 0.0004998). In contrast, the distribution of SNVs in a scrambled version of the RCGTG motif (RGCGT) was similar regardless the location of the motif in HBR (Figure 7B). Altogether these results suggest that variants affecting functional RCGTG motifs are under negative selection, which decreases the frequency of polymorphisms at these loci.

DISCUSSION

Genome-wide association studies indicate that a large fraction of the SNVs linked to specific phenotypic traits, including increased risk of particular diseases, map to the non-coding portion of the genome. Moreover, disease-associated variants are not randomly distributed throughout the genome but tend to cluster in regulatory regions such as promoters and enhancers (61). However, even with the aid of the functional list of elements provided by the ENCODE project (49), the interpretation of the significance of variants in non-coding regions remains challenging. Hence, it is of great interest to identify variants that have a functional effect and the molecular mechanism behind them.

Herein, we report a curated list of genome-wide HIF binding regions and characterize the variability of the RCGTG motifs within them. Although this study is far from being a comprehensive description of the genetic variability of all functional HREs in the human genome, to our knowledge, this is the first attempt to systematically analyze the inter-individual variability in the transcriptional response to hypoxia.

One limitation of our study is that, due to cost-efficiency and practical reasons, we restricted our study to a subset of all potential HBR and to frequent variants disrupting RCGTG motifs within them. These restrictions narrowed the focus of our analysis to 14 variants. However, we found many other polymorphisms that could potentially affect the

---

ing (Supplementary Table SIV). The induction of *EGLN3* (right graph) and *EGFR* (left graph) genes by hypoxia (1% oxygen for 16 h) was determined by quantitative PCR in each cell clone and represented as the median induction of the gene in several (between 2 and 4) independent experiments, normalized to the median induction of the wild-type clones. Each dot represents an independent clone and shape represents their genotype. Clones that contain the reference sequence in the three alleles of the HRE present in HEK293T cells are shown as circles, clones with at least a single functional copy are shown as squares and clones with non-functional copies are shown as triangles. The specific genotype is represented using different shapes as indicated in the figure legend. Note that genotypes in the right graph refer to the HRE in the *EGLN3* locus and those in the right graph to the HRE in *EGFR*. The solid gray line represent the mean of the normalized induction of wild-type clones and the dotted lined the 95% confidence interval of the mean. The line at 0.5 marks a normalized induction half of the expected for the wild-type clones. The difference between the mean induction of *EGLN3* in wild-type clones and those with an impaired copy of the EGLN3 HRE was significant (student's t-test: $t_{4.7} = 5.9$, $P = 0.0025$). the mean induction of *EGFR* in EGFR in clones with all HREs copies disrupted was significantly lower than that observed for clones with all functional alleles (ANOVA $F_{2,1}4 = 4.987$, $P < 0.05$, pos-hoc Tukey test $P < 0.05$). (**E**) The effect of hypoxia on the transcriptional activity of the *EGFR* enhancer region was assessed upon transfection of HeLa cells with constructs in which the luciferase was under the control of the wild-type sequence ('WT'), a sequence containing variant chr7:55254186:A ('SNP') or a mutant sequence with a truncated HRE ('MUT'). The graph represents the normalized luciferase activity in hypoxic samples expressed as fold over the activity obtained in normoxic conditions. Each symbol represent the results obtained in an independent experiment. The differences between groups were statistically significant (ANOVA $F_{3,11} = 55.1$, $P < 0.001$) and the mean induction of the WT group was significantly different to that of SNP and MUT (adjusted $P < 0.001$) in a *posteriori* Tukey test.
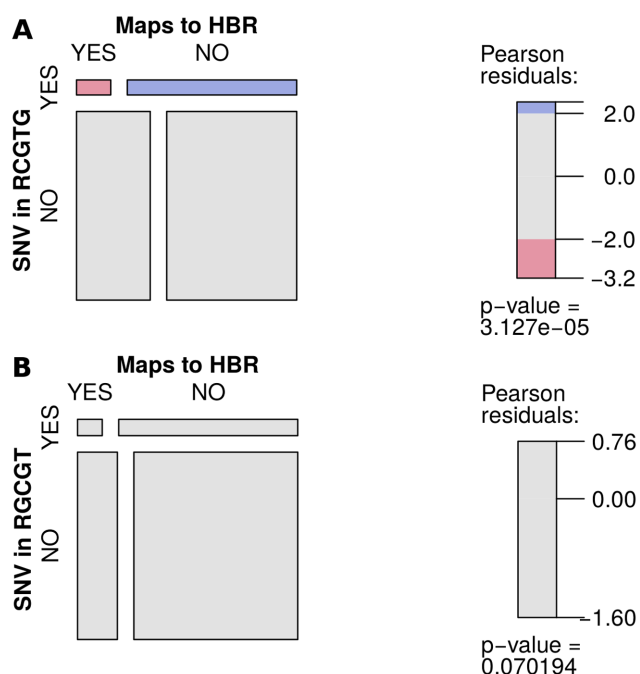
Figure 7. Association between presence of variants and HIF binding. The whole set of RCGTG (HRE, panel **A**) or RGCGT (scrambled version of the HRE, panel **B**) motifs found in the human genome were classified according to their location in HBR ('YES'/'NO' in the categorical variable 'Maps to HBR') and the number of motifs in each location harboring SNV recorded. The figure represents the number of motifs in each combination of categories (presence/absence of variants and location inside/outside HBRs). Since the actual number of motifs outside HBRs is very large, to facilitate visualization, we represent a random sample of 1000 of these motifs. The colors represent the divergence of the actual numbers from the expected counts under the null hypothesis of independence between both variables measured as the Pearson residuals.

response to hypoxia. In this regard, genomic variants in HBR could alter HIF binding without disrupting the sequence of RCGTG motifs, as was described for the regulation of cyclin D by EPAS1 in renal cancer (22). These polymorphisms could exert their effect by, for example, altering the binding of additional transcription factors in the proximity of an HRE that cooperate with HIF in the transcriptional regulation (62). In addition, we also found some variants that result in the generation of novel RCGTG motifs within the HBR. These variants might increase the strength of the response to hypoxia, as was demonstrated for a polymorphism in the chicken PGK1 locus (63). Finally, we also found polymorphisms that, while keeping a functional motif, might alter the affinity for HIF such as the change of ACGTG for GCGTG. All these alterations and their effect on hypoxia-induced transcription will be addressed in future works.

Limitations aside, we attempted the functional characterization of all the common variants found in our sequence analysis that disrupted RCGTGs motifs (Figure 4). These experiments revealed that 8 out of 10 analyzed regions were responsive to hypoxia (log ratio hypoxia to normoxia of the reference allele was significantly different of zero, single sample *t*-test, $P < 0.05$,). The lack of response of in some of the constructs (regions associated to *PFKP, PPME1*

and *RNMT*) is puzzling because we defined the regulatory regions based on multiple independent experimental evidences of HIF binding (see Supplementary Table SI). A possible explanation for this result is that some of the regions in our list of HBRs could represent chromatin indirectly recruited during ChIP by tethering rather than being directly bound by HIF, as it has been reported for many other transcription factors (64). Alternatively, technical issues might explain these results; for example, the selected region might lack some cis-regulatory elements required for a proper transcriptional response to hypoxia or the cells used for the reporter assays might lack a necessary transacting factor. Another interesting case is the region associated to *ATP9A* as it represent a HIF-bound regulatory region whose transcriptional activity is repressed by hypoxia. Global analysis (21,51,53) have failed to demonstrate a significant association between HIF binding and gene downregulation. However, it is possible that, in unusual cases, direct HIF binding could lead to transcriptional repression as suggested for the regulation of *CAD and AFP* genes by hypoxia (65,66). Nevertheless, most of the cloned regions responded to hypoxia albeit with different strengths.

Remarkably, the comparison of the activity of reporter constructs carrying the reference and alternate variants showed that variants rs6593210, rs1009329 and rs150921338 significantly affected the transcriptional response to hypoxia, supporting that genetic variability results in inter-individual differences in the hypoxia-induced gene expression profile. It is also noteworthy that, according to our data, a single nucleotide change regardless the position affected in the RCGTG motif results in a complete lost of function. This result might be of help for the interpretation of the effect of variants found within functional HREs.

In spite of these positive results, we found that, in most of the cases, the variants altering RCGTG motifs had no effect or only a partial effect on the response to hypoxia as in the case of *PGK1*. Moreover, detailed analysis of the variant chr7:55254186G>A (rs6593210) within the *EGFR* locus revealed that, although the alternate allele resulted in a complete inhibition of the transcriptional activity of the isolated enhancer, when assayed in the native genomic context, the variant showed a partial effect. These results are in agreement with our observation of decreased frequencies of polymorphisms altering the (wild-type) sequence of HREs in comparison to non-functional RCGTG motifs and provide further support to the notion of evolutionary constraint of functional HREs due to negative selection. This conclusion is in agreement with the reduced variability observed for other transcription factors binding sites (49,67).

In view of these results, it is not surprising the lack of evidence for association between the studied variants and phenotypes in COPD and OSAS patients (Figure 3), as most of them do not affect functional HREs (Figure 4). In the case of the variants that affect the response to hypoxia, we cannot rule out that the reduced sample size in our study and/or a small effect size prevented us from finding some underlying association. In this regard, rs6593210 affects a functional HRE (Figures 4 and 6E), albeit redundant, and the alternate allele reduces de induction of *EGF*R in response to hypoxia (Figure 6D). Since activation of the epi-

dermal growth factor receptor has been implicated in the overproduction of mucus, one of the hallmarks of COPD (68–70), the lack of evidence of association between variant rs6593210 and COPD progression markers should be taken with caution and future studies should address this particular question. In addition, the EGFR pathway is critical for the progression of multiple neoplasias and, intriguingly, COPD patients have increased risk of developing lung cancer. Thus, given the results presented herein, it would be interesting to study the effect of rs6593210 variant in the progression of cancer in situations where HIF is constitutively active such as ccRCC and lung cancer in COPD patients.

On the other hand, as indicated above, further work is required to assess whether variants outside the RCGTG motifs could alter the response to hypoxia and contribute to the progression of these diseases. Finally, our study revealed a large number of low-frequency variants that alter the sequence of RCGTG motifs. It would be interesting to investigate whether these variants affect functional HREs and, if so, study their potential impact on the phenotype of the patients.

In summary, in this work we defined a list of HIF-bound regions supported by several robust experimental evidences and studied the genetic variability affecting the RCGTG motifs within them. Although our analysis suggests that functional HREs tend to be invariant, we found three SNPs (rs1009329, rs6593210 and rs150921338) that significantly attenuate the transcriptional response to hypoxia. Thus, although disfavored by selection, there exists some subtle variability in the inter-individual response to hypoxia whose effect on disease awaits further work.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

Conflict of interest statement. None declared.

REFERENCES

1. Semenza,G.L. (2014) Oxygen sensing, hypoxia-inducible factors, and disease pathophysiology. *Annu. Rev. Pathol.*, **9**, 47–71.
2. 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.
3. Wang,G.L., Jiang,B.H., Rue,E.A. and Semenza,G.L. (1995) Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O2 tension. *Proc. Natl. Acad. Sci. U.S.A.*, **92**, 5510–5514.
4. Ivan,M., Kondo,K., Yang,H., Kim,W., Valiando,J., Ohh,M., Salic,A., Asara,J.M., Lane,W.S. and Kaelin,W.G. (2001) HIFalpha targeted for VHL-mediated destruction by proline hydroxylation: implications for O2 sensing. *Science*, **292**, 464–468.
5. Jaakkola,P., Mole,D.R., Tian,Y.M., Wilson,M.I., Gielbert,J., Gaskell,S.J., von Kriegsheim,A., Hebestreit,H.F., Mukherji,M., Schofield,C.J. *et al.* (2001) Targeting of HIF-alpha to the von Hippel-Lindau ubiquitylation complex by O2-regulated prolyl hydroxylation. *Science*, **292**, 468–472.
6. Bruick,R.K. and McKnight,S.L. (2001) A conserved family of prolyl-4-hydroxylases that modify HIF. *Science*, **294**, 1337–1340.
7. 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.
8. 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.
9. Lando,D., Peet,D.J., Gorman,J.J., Whelan,D.a, Whitelaw,M.L. 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.
10. Hirsilä,M., Koivunen,P., Günzler,V., Kivirikko,K.I. and Myllyharju,J. (2003) Characterization of the human prolyl 4-hydroxylases that modify the hypoxia-inducible factor. *J. Biol. Chem.*, **278**, 30772–30780.
11. Lee,F.S. (2008) Genetic causes of erythrocytosis and the oxygen-sensing pathway. *Blood Rev.*, **22**, 321–332.
12. Bigham,A.W. and Lee,F.S. (2014) Human high-altitude adaptation: forward genetics meets the HIF pathway. *Genes Dev.*, **28**, 2189–2204.
13. Latif,F., Tory,K., Gnarra,J., Yao,M., Duh,F.M., Orcutt,M.L., Stackhouse,T., Kuzmin,I., Modi,W. and Geil,L. (1993) Identification of the von Hippel-Lindau disease tumor suppressor gene. *Science*, **260**, 1317–1320.
14. Ang,S.O., Chen,H., Hirota,K., Gordeuk,V.R., Jelinek,J., Guan,Y., Liu,E., Sergueeva,A.I., Miasnikova,G.Y., Mole,D. *et al.* (2002) Disruption of oxygen homeostasis underlies congenital Chuvash polycythemia. *Nat. Genet.*, **32**, 614–621.
15. Yang,C., Zhuang,Z., Fliedner,S.M.J., Shankavaram,U., Sun,M.G., Bullova,P., Zhu,R., Elkahloun,A.G., Kourlas,P.J., Merino,M. *et al.* (2015) Germ-line PHD1 and PHD2 mutations detected in patients with pheochromocytoma/paraganglioma-polycythemia. *J. Mol. Med. (Berl).*, **93**, 93–104.
16. Percy,M.J., Beer,P.A., Campbell,G., Dekker,A.W., Green,A.R., Oscier,D., Rainey,M.G., van Wijk,R., Wood,M., Lappin,T.R.J. *et al.*

(2008) Novel exon 12 mutations in the HIF2A gene associated with erythrocytosis. *Blood*, **111**, 5400–5402.

17. Percy,M.J., Furlow,P.W., Lucas,G.S., Li,X., Lappin,T.R.J., McMullin,M.F. and Lee,F.S. (2008) A gain-of-function mutation in the HIF2A gene in familial erythrocytosis. *N. Engl. J. Med.*, **358**, 162–168.

18. Ladroue,C., Carcenac,R., Leporrier,M., Gad,S., Le Hello,C., Galateau-Salle,F., Feunteun,J., Pouysségur,J., Richard,S. and Gardie,B. (2008) PHD2 mutation and congenital erythrocytosis with paraganglioma. *N. Engl. J. Med.*, **359**, 2685–2692.

19. Yi,X., Liang,Y., Huerta-Sanchez,E., Jin,X., Cuo,Z.X.P., Pool,J.E., Xu,X., Jiang,H., Vinckenbosch,N., Korneliussen,T.S. *et al.* (2010) Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*, **329**, 75–78.

20. Buroker,N.E., Ning,X.-H., Zhou,Z.-N., Li,K., Cen,W.-J., Wu,X.-F., Zhu,W.-Z., Scott,C.R. and Chen,S.-H. (2012) EPAS1 and EGLN1 associations with high altitude sickness in Han and Tibetan Chinese at the Qinghai-Tibetan Plateau. *Blood Cells. Mol. Dis.*, **49**, 67–73.

21. Ortiz-Barahona,A., Villar,D., Pescador,N., Amigo,J. and del Peso,L. (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**, 2332–2345.

22. Schödel,J., Bardella,C., Sciesielski,L.K., Brown,J.M., Pugh,C.W., Buckle,V., Tomlinson,I.P., Ratcliffe,P.J. and Mole,D.R. (2012) Common genetic variants at the 11q13.3 renal cancer susceptibility locus influence binding of HIF to an enhancer of cyclin D1 expression. *Nat. Genet.*, **44**, 420–425.

23. Miravitlles,M., Soriano,J.B., García-Río,F., Muñoz,L., Duran-Tauleria,E., Sanchez,G., Sobradillo,V. and Ancochea,J. (2009) Prevalence of COPD in Spain: impact of undiagnosed COPD on quality of life and daily life activities. *Thorax*, **64**, 863–868.

24. Peppard,P.E., Young,T., Barnet,J.H., Palta,M., Hagen,E.W. and Hla,K.M. (2013) Increased prevalence of sleep-disordered breathing in adults. *Am. J. Epidemiol.*, **177**, 1006–1014.

25. Decramer,M., Janssens,W. and Miravitlles,M. (2012) Chronic obstructive pulmonary disease. *Lancet*, **379**, 1341–1351.

26. Lévy,P., Kohler,M., McNicholas,W.T., Barbé,F., McEvoy,R.D., Somers,V.K., Lavie,L. and Pépin,J.-L. (2015) Obstructive sleep apnoea syndrome. *Nat. Rev. Dis. Prim.*, **1**, 15015.

27. Jiang,H., Zhu,Y., Xu,H., Sun,Y. and Li,Q. (2010) Activation of hypoxia-inducible factor-1alpha via nuclear factor-kappa B in rats with chronic obstructive pulmonary disease. *Acta Biochim. Biophys. Sin. (Shanghai)*, **42**, 483–488.

28. Belaidi,E., Joyeux-Faure,M., Ribuot,C., Launois,S.H., Levy,P. and Godin-Ribuot,D. (2009) Major role for hypoxia inducible factor-1 and the endothelin system in promoting myocardial infarction and hypertension in an animal model of obstructive sleep apnea. *J. Am. Coll. Cardiol.*, **53**, 1309–1317.

29. Yuan,G., Nanduri,J., Khan,S., Semenza,G.L. and Prabhakar,N.R. (2008) Induction of HIF-1alpha expression by intermittent hypoxia: involvement of NADPH oxidase, Ca2+ signaling, prolyl hydroxylases, and mTOR. *J. Cell. Physiol.*, **217**, 674–685.

30. Yu,A.Y., Shimoda,L.A., Iyer,N. V., Huso,D.L., Sun,X., McWilliams,R., Beaty,T., Sham,J.S., Wiener,C.M., Sylvester,J.T. *et al.* (1999) Impaired physiological responses to chronic hypoxia in mice partially deficient for hypoxia-inducible factor 1alpha. *J. Clin. Invest.*, **103**, 691–696.

31. Brusselmans,K., Compernolle,V., Tjwa,M., Wiesener,M.S., Maxwell,P.H., Collen,D. and Carmeliet,P. (2003) Heterozygous deficiency of hypoxia-inducible factor-2alpha protects mice against pulmonary hypertension and right ventricular dysfunction during prolonged hypoxia. *J. Clin. Invest.*, **111**, 1519–1527.

32. Hickey,M.M., Richardson,T., Wang,T., Mosqueira,M., Arguiri,E., Yu,H., Yu,Q., Solomides,C.C., Morrisey,E.E., Khurana,T.S. *et al.* (2010) The von Hippel–Lindau Chuvash mutation promotes pulmonary hypertension and fibrosis in mice. *J. Clin. Invest.*, **120**, 827–839.

33. Peng,Y.-J., Yuan,G., Ramakrishnan,D., Sharma,S.D., Bosch-Marce,M., Kumar,G.K., Semenza,G.L. and Prabhakar,N.R. (2006) Heterozygous HIF-1α deficiency impairs carotid body-mediated systemic responses and reactive oxygen species generation in mice exposed to intermittent hypoxia. *J. Physiol.*, **577**, 705–716.

34. Gale,D.P., Harten,S.K., Reid,C.D.L., Tuddenham,E.G.D. and Maxwell,P.H. (2008) Autosomal dominant erythrocytosis and pulmonary arterial hypertension associated with an activating HIF2 mutation. *Blood*, **112**, 919–921.

35. Tan,Q., Kerestes,H., Percy,M.J., Pietrofesa,R., Chen,L., Khurana,T.S., Christofidou-Solomidou,M., Lappin,T.R.J. and Lee,F.S. (2013) Erythrocytosis and pulmonary hypertension in a mouse model of human HIF2A gain of function mutation. *J. Biol. Chem.*, **288**, 17134–17144.

36. Yoo,S., Takikawa,S., Geraghty,P., Argmann,C., Campbell,J., Lin,L., Huang,T., Tu,Z., Feronjy,R., Spira,A. *et al.* (2015) Integrative analysis of DNA methylation and gene expression data identifies EPAS1 as a key regulator of COPD. *PLoS Genet.*, **11**, e1004898.

37. Miravitlles,M., Soler-Cataluña,J.J., Calle,M., Molina,J., Almagro,P., Quintano,J.A., Riesco,J.A., Trigueros,J.A., Piñera,P., Simón,A. *et al.* (2014) Spanish guideline for COPD (GesEPOC). Update 2014. *Arch. Bronconeumol.*, **50**(Suppl. 1), 1–16.

38. Celli,B.R., Cote,C.G., Marin,J.M., Casanova,C., de Oca,M., mendez,R.A., Pinto-Plata,V. and Cabral,H.J. (2004) The body mass index, airflow obstruction, dyspnea and exercise capacity index in chronic obstructive pulmonary disease. *N. Engl. J. Med.*, **350**, 1005–1012.

39. Grundy,S.M., Brewer,H.B. Jr, Cleeman,J.I., Smith,S.C. Jr, Lenfant,C., American Heart Association, National Heart Lung and Blood Institute (2004) Definition of metabolic syndrome: Report of the National Heart, Lung, and Blood Institute/American Heart Association conference on scientific issues related to definition. *Circulation*, **109**, 433–438.

40. Alpert,J.S., Thygesen,K., Antman,E. and Bassand,J.P. (2000) Myocardial infarction redefined–a consensus document of The Joint European Society of Cardiology/American College of Cardiology Committee for the redefinition of myocardial infarction. *J. Am. Coll. Cardiol.*, **36**, 959–969.

41. Divo,M., Cote,C., de Torres,J.P., Casanova,C., Marin,J.M., Pinto-Plata,V., Zulueta,J., Cabrera,C., Zagaceta,J., Hunninghake,G. *et al.* (2012) Comorbidities and risk of mortality in patients with chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.*, **186**, 155–161.

42. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.

43. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.

44. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G., Durbin,R. and 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

45. Robinson,J.T., Thorvaldsdóttir,H., Winckler,W., Guttman,M., Lander,E.S., Getz,G. and Mesirov,J.P. (2011) Integrative genomics viewer. *Nat. Biotechnol.*, **29**, 24–26.

46. Lawrence,M., Huber,W., Pages,H., Aboyoun,P., Carlson,M., Gentleman,R., Morgan,M.T. and Carey,V.J. (2013) Software for computing and annotating genomic ranges. *PLoS Comput. Biol.*, **9**, e1003118.

47. Dölken,L., Ruzsics,Z., Rädle,B., Friedel,C.C., Zimmer,R., Mages,J., Hoffmann,R., Dickinson,P., Forster,T., Ghazal,P. *et al.* (2008) High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA*, **14**, 1959–1972.

48. Schwanhäusser,B., Busse,D., Li,N., Dittmar,G., Schuchhardt,J., Wolf,J., Chen,W. and Selbach,M. (2011) Global quantification of mammalian gene expression control. *Nature*, **473**, 337–342.

49. ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.

50. 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 and HIF-2 DNA binding with expression profiling of hypoxia-inducible transcripts. *J. Biol. Chem.*, **284**, 16767–16775.

51. 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. U.S.A.*, **106**, 4260–4265.

52. Xia,X. and Kung,A.L. (2009) Preferential binding of HIF-1 to transcriptionally active loci determines cell-type specific response to hypoxia. *Genome Biol.*, **10**, R113.

53. Schödel,J., Oikonomopoulos,S., Ragoussis,J., Pugh,C.W., Ratcliffe,P.J. and Mole,D.R. (2011) High-resolution genome-wide mapping of HIF-binding sites by ChIP-seq. *Blood*, **117**, e207–e217.

54. Semenza,G.L., Roth,P.H., Fang,H.M. and Wang,G.L. (1994) Transcriptional regulation of genes encoding glycolytic enzymes by hypoxia-inducible factor 1. *J. Biol. Chem.*, **269**, 23757–23763.

55. Franovic,A., Gunaratnam,L., Smith,K., Robert,I., Patten,D. and Lee,S. (2007) Translational up-regulation of the EGFR by tumor hypoxia provides a nonmutational explanation for its overexpression in human cancer. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 13092–13097.

56. Tiana,M., Villar,D., Pérez-Guijarro,E., Gómez-Maldonado,L., Moltó,E., Fernández-Miñán,A., Gómez-Skarmeta,J.L.J.L., Montoliu,L. and Del Peso,L. (2012) A role for insulator elements in the regulation of gene expression response to hypoxia. *Nucleic Acids Res.*, **40**, 1916–1927.

57. Dölken,L., Ruzsics,Z., Rädle,B., Friedel,C.C., Zimmer,R., Mages,J., Hoffmann,R., Dickinson,P., Forster,T., Ghazal,P. *et al.* (2008) High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA*, **14**, 1959–1972.

58. Sanjana,N.E., Cong,L., Zhou,Y., Cunniff,M.M., Feng,G. and Zhang,F. (2012) A transcription activator-like effector toolbox for genome engineering. *Nat. Protoc.*, **7**, 171–192.

59. Pescador,N., Cuevas,Y., Naranjo,S., Alcaide,M., Villar,D., Landázuri,M.O. and Del Peso,L. (2005) Identification of a functional hypoxia-responsive element that regulates the expression of the egl nine homologue 3 (egln3/phd3) gene. *Biochem. J.*, **390**, 189–197.

60. Lin,Y.-C., Boone,M., Meuris,L., Lemmens,I., Van Roy,N., Soete,A., Reumers,J., Moisse,M., Plaisance,S., Drmanac,R. *et al.* (2014) Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nat. Commun.*, **5**, 4767.

61. Zhang,F. and Lupski,J.R. (2015) Non-coding genetic variants in human disease. *Hum. Mol. Genet.*, **24**, R102–R110.

62. Villar,D., Ortiz-Barahona,A., Gómez-Maldonado,L., Pescador,N., Sánchez-Cabo,F., Hackl,H., Rodriguez,B.A., Trajanoski,Z., Dopazo,A., Huang,T.H. *et al.* (2012) Cooperativity of Stress-Responsive Transcription Factors in Core Hypoxia-Inducible Factor Binding Regions. *PLoS One*, **7**, e45708.

63. Wang,C., Yuan,C., Zhang,L., Wu,C. and Li,N. (2007) Differential gene expression of phosphoglyceric kinase (PGK) and hypoxic adaptation in chicken. *Sci. China Ser. C Life Sci.*, **50**, 335–342.

64. Neph,S., Vierstra,J., Stergachis,A.B., Reynolds,A.P., Haugen,E., Vernot,B., Thurman,R.E., John,S., Sandstrom,R., Johnson,A.K. *et al.* (2012) An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature*, **489**, 83–90.

65. Chen,K.-F., Lai,Y.-Y., Sun,H.S. and Tsai,S.-J. (2005) Transcriptional repression of human cad gene by hypoxia inducible factor-1alpha. *Nucleic Acids Res.*, **33**, 5190–5198.

66. Mazure,N.M., Chauvet,C., Bois-Joyeux,B., Bernard,M.-A., Nacer-Chérif,H. and Danan,J.-L. (2002) Repression of alpha-fetoprotein gene expression under hypoxic conditions in human hepatoma cells: characterization of a negative hypoxia response element that mediates opposite effects of hypoxia inducible factor-1 and c-Myc. *Cancer Res.*, **62**, 1158–1165.

67. Moyerbrailean,G.A., Kalita,C.A., Harvey,C.T., Wen,X., Luca,F. and Pique-Regi,R. (2016) Which genetics variants in DNase-Seq footprints are more likely to alter binding? *PLOS Genet.*, **12**, e1005875.

68. Yu,H., Li,Q., Kolosov,V.P., Perelman,J.M. and Zhou,X. (2012) Regulation of cigarette smoke-mediated mucin expression by hypoxia-inducible factor-1α via epidermal growth factor receptor-mediated signaling pathways. *J. Appl. Toxicol.*, **32**, 282–292.

69. Takeyama,K., Jung,B., Shim,J.J., Burgel,P.R., Dao-Pick,T., Ueki,I.F., Protin,U., Kroschel,P. and Nadel,J.a (2001) Activation of epidermal growth factor receptors is responsible for mucin synthesis induced by cigarette smoke. *Am. J. Physiol. Lung Cell. Mol. Physiol.*, **280**, L165–L172.

70. Kim,S., Shim,J.J., Burgel,P.-R., Ueki,I.F., Dao-Pick,T., Tam,D.C.-W. and Nadel,J.A. (2002) IL-13-induced Clara cell secretory protein expression in airway epithelium: role of EGFR signaling pathway. *Am. J. Physiol. Lung Cell. Mol. Physiol.*, **283**, L67–L75.

71. Yu,G., Wang,L.-G., Han,Y. and He,Q.-Y. (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*, **16**, 284–287.

72. Kent,W.J., Sugnet,C.W., Furey,T.S., Roskin,K.M., Pringle,T.H., Zahler,A.M. and Haussler,D. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.