Genome-wide mapping of the hypoxic response

A thesis submitted for the degree of Doctor of Philosophy

Michaelmas Term 2012

Dr. med. Johannes Schödel

Brasenose College



Abstract

Genome-wide mapping of the hypoxic response

Dr. med. Johannes Schödel, Brasenose College
Thesis submitted for the degree of Doctor of Philosophy
Centre for Cellular and Molecular Physiology, University of Oxford
Michaelmas Term 2012

Hypoxia regulates many hundreds of genes with important roles in ischemic and neoplastic disorders. Central to this response are the hypoxia inducible transcription factors (HIF). This work aimed to better understand the direct transcriptional response to HIF by mapping HIF-binding sites across the genome using chromatin immunoprecipitation coupled to high-throughput sequencing (ChIP-seq). ChIP-seq for HIF in MCF-7 breast cancer cells under hypoxic conditions revealed more than 400 high-stringency HIF-binding sites genome-wide. Each member of the HIF heterodimer was present with near complete concordance. Binding of the two principle isoforms revealed a high degree of overlap with no differences in the DNA-binding motif. HIF-binding was associated with upregulation, but not downregulation of genes indicating that it functions as a transcriptional activator but not as a repressor. HIF-binding occurred preferentially at gene promoters, but was also present at promoter-distant sites, which were also associated with gene regulation, implicating long-range interactions in hypoxic gene activation. HIF-binding was associated with markers of open chromatin and active enhancers that were present in normoxia, indicating that HIF-binding sites are already "prepared" to bind HIF before the hypoxic stimulus. Analysis of normoxic and hypoxic RNA pol2 and H3K4me3 signals revealed distinctive hypoxia-inducible changes unique to HIF-binding genes. Comparable numbers of HIF-binding sites were observed in a second cell line (von Hippel-Lindau defective 786-O renal cancer cells) as in MCF-7 breast cancer cells, although approximately 65% were unique to 786-O cells. These unique sites were more frequently promoter-distant. Correlation with expression analyses from renal tumours indicated that many HIF-binding genes were upregulated in renal cancer. One such RCC unique promoter-distant HIF-binding site was identified at an intergenic locus on chromosome 11q13.3 that has been associated with renal cancer in Genome-Wide Association Studies. The HIF-binding site was in high linkage disequilibrium with the disease associated SNP and had the epigenetic hallmarks of an enhancer. Analysis of pan-genomic expression analyses identified the cell-cycle regulator cyclin D1 as highly HIF-regulated, and a physical association between the HIF-binding site and the *CCND1* promoter could be determined. Furthermore, in a renal cancer cell line heterozygous at this locus, the RCC-protective allele disrupted HIF-binding leading to an allelic imbalance in cyclin D1 expression.

Table of Contents

1 CHAPTER 1: Introduction 1
1.1 The hypoxia inducible transcription factor (HIF) pathway 2
1.1.1 Atmospheric oxygen levels and evolution of life 2
1.1.2 The oxygen sensing pathway 3
1.1.3 Oxygen levels and HIF in humans 6
1.1.4 PHD-VHL-HIF pathway in humans 7
1.1.5 Effects of HIF on gene expression 9
1.2 HIF in cancer 12
1.3 VHL syndrome and renal cancer 13
1.4 The divergent roles of HIF-1 and HIF-2 in renal cancer 15
1.5 Mechanisms of differential functions of HIF-1α and HIF-2α 16
1.6 Genetic alterations in renal cancer 19
1.6.1 Rare familial syndromes 19
1.6.2 Genome-wide approaches in sporadic RCC 20
1.7 Profiling of the hypoxic response and defining HIF-binding sites 23
1.8 Aim of this study 27
2 CHAPTER 2: Genome-wide HIF-binding in MCF-7 cells using ChIP-seq 29
2.1 Introduction 30
2.2 Results 32
2.2.1 ChIP and ChIP-seq for HIF-α subunits 32
2.2.2 Genome-wide HIF-1α and HIF-2α binding 34
2.2.3 HIF-1β binding in MCF-7 cells 37
2.2.4 Distribution of HIF-binding in the vicinity of transcriptional start sites 41
2.2.5 Distribution of HIF-binding across the genome 43
2.2.6 Pathway analysis 44
2.2.7 Analysis of transcription factor binding site motifs at HIF DNA-binding sites 45
2.2.8 Correlation of HIF-binding with DNAse1 hypersensitivity 48
2.2.9 Correlation of HIF-binding sites with CpG islands 52
2.2.10 Normoxic HIF-1β binding 54
2.2.11 HIF-binding and gene regulation 55

2.3 Discussion 67
2.3.1 HIF-α and HIF-β binding in MCF-7 cells 67
2.3.2 Definition of HIF-binding sites 68
2.3.3 Genetic and epigenetic environment at HIF-binding sites 69
2.3.4 HIF-binding and gene regulation 71
3 CHAPTER 3: Genome-wide analysis of histone 3 lysine 4 methylation and RNA pol2 in MCF-7 cells under normoxic and hypoxic conditions 73
3.1 Introduction 74
3.2 Results 78
3.2.1 Changes of histone modifications under hypoxic conditions 78
3.2.2 Changes of histone modifications and RNA pol2 binding under hypoxic conditions at selected gene loci 79
3.2.3 Distribution of histone marks and RNA pol2 under normoxic conditions 81
3.2.4 ChIP-seq for H3K4me1, H3K4me3 and RNA pol2 under normoxic and hypoxic conditions 84
3.2.4.1 Changes of RNA pol2 and histone marks at promoters and gene bodies under hypoxic conditions 84
3.2.4.1.1 RNA pol2 87
3.2.4.1.2 H3K4me3 88
3.2.4.1.3 H3K4me1 90
3.2.4.2 Changes of RNA pol2 and histone marks at promoters and gene bodies of HIF-binding and non HIF-binding genes upregulated by hypoxia 91
3.2.4.3 Changes of RNA pol2 and histone marks under hypoxic conditions at HIF-binding sites distant from promoters 94
3.2.5 Detection of non-conventional HIF-binding sites: HIF-1α anti-sense transcript ... 100
3.2.6 Detection of non-conventional HIF-binding sites: a previously unannotated intergenic transcript - "Nici" 103
3.3 Discussion 108
3.3.1 Regulation of histone modifications and RNA pol2 in hypoxia 109
3.3.2 Mechanisms involved in increased methylation of histone 3 lysine 4 in hypoxia . 110
3.3.3 Regulation of transcript abundance in hypoxia 114
3.3.4 Non-conventional HIF-binding at remote loci: potential functions of HIF-1α antisense and "Nici" 115
4 CHAPTER 4: Genome-wide HIF-binding in 786-O renal cancer cells 118
4.1 Introduction 119
4.2 Results 121

4.2.1 HIF-binding in different renal cancer cell lines 121
4.2.2 HIF-binding in 786-O RCC cells 122
4.2.3 ChIP-seq for HIF-2α and HIF-1β in 786-O cells 124
4.2.4 Overlap of HIF-binding sites between MCF-7 and 786-O cells 125
4.2.5 Motif analysis in 786-O 126
4.2.6 Distribution of HIF-binding in 786-O cells compared to MCF-7 cells 128
4.2.7 Gene regulation by HIF-2α in 786-O cells 129
4.2.8 Association between HIF-2 binding in 786-O cells and genes deregulated in renal cancer tissue 134
4.3 Discussion 136
5 CHAPTER 5: Identification of a HIF-binding enhancer of cyclin D1 expression specific for VHL-defective renal cancer 139
5.1 Introduction 140
5.2 Results 142
5.2.1 HIF-binding to an intergenic remote region on 11q13.3 142
5.2.2 Cell type specific HIF-binding and accessibility 144
5.2.3 The 11q13.3 HIF-binding site exhibits the hallmarks of an active enhancer 151
5.2.4 The 11q13.3 H locus is the main HIF-binding regulatory site 154
5.2.5 HIF regulates cyclin D1 and the enhancer loops to the *CCND1* promoter 155
5.2.6 Genetic evidence that the enhancer is relevant in renal cancer 161
5.2.7 11q13.3 polymorphisms influence HIF-binding to the *CCND1* enhancer 163
5.2.8 Polymorphisms at the enhancer modulate cyclin D1 expression 167
5.3 Discussion 169
5.3.1 Genetic variants at remote regulatory elements 169
5.3.2 Association of enhancer activity with loss of functional VHL 171
5.3.3 Implications for further studies 172
6 CHAPTER 6: Final Discussion 174
7 CHAPTER 7: Material and Methods 178
7.1 Cell lines 179
7.2 Cell Culture 180
7.3 DNA-techniques 180
7.3.1 General DNA techniques 180
7.3.1.1 Isolation of genomic DNA 180
7.3.1.2 Quantification of DNA concentration 181

7.3.1.3 Polymerase chain reaction (PCR) amplification of DNA 182
7.3.1.4 Restriction digest of DNA 183
7.3.1.5 Gel electrophoresis 183
7.3.1.6 Purification of amplified or digested DNA 183
7.3.1.7 Ligation of DNA fragments 184
7.3.1.8 Generation of vector constructs for transfection 184
7.3.1.8.1 Bacterial strains, transformation and isolation of constructs 185
7.3.1.8.2 DNA sequencing 186
7.3.1.8.3 Cloning of specific constructs 186
7.3.1.8.4 Site directed mutagenesis (SDMG) 187
7.3.1.8.5 Transfection of plasmid DNA 188
7.3.1.8.6 Reporter assays 189
7.4 Protein Techniques 190
7.4.1 Protein lysates 190
7.4.2 Immunoblotting 191
7.5 RNA techniques 192
7.5.1 RNA isolation from cultured cells 192
7.5.2 Control of RNA quality 193
7.5.3 Generation of complementary DNA 193
7.5.4 Expression and enrichment analysis using Taqman® or Sybr green® assays 194
7.5.5 Settings for qPCR experiments 195
7.5.6 Data analysis 196
7.5.7 Microarray analysis 197
7.5.8 Transfection of silencing RNA (siRNA) 197
7.6 Specific DNA techniques 199
7.6.1 Preparation of BAC, FOSMID and Image clones 199
7.6.2 2D and 3D FISH experiments 200
7.6.3 Chromatin immunoprecipitation (ChIP) 201
7.6.3.1 Quality control for ChIP 205
7.6.3.2 ChIP-sequencing 205
7.6.3.3 Data and Data analysis tools 208
7.6.3.4 Independent confirmation of enrichment by real-time qPCR 208
7.6.4 Formaldehyde assisted isolation of regulatory elements (FAIRE) 209
7.6.5 Chromatin conformation capture (3C) assay 211

7.6.5.1 Quality control for 3C assays 214
7.6.5.2 qPCR analysis of 3C templates 214
7.6.6 Allele specific assays 215
7.7 Antibodies 216
7.8 Primers and Taqman® assays 217
7.8.1 Primers 217
7.8.2 Taqman® assays 222
7.9 Solutions and buffers 223
7.9.1 Urea- sodium dodecyl sulphate (SDS) buffer 223
7.9.2 Western blotting running buffer (10x) 223
7.9.3 Western blotting transfer buffer (20X) 224
7.9.4 SDS-PAGE resolving gel buffer 224
7.9.5 SDS-PAGE stacking gel buffer 224
7.9.6 Phosphate buffered saline (PBS) 224
7.9.7 PBS-Tween (PBS-T) 225
7.9.8 5% Milk in PBS-T 225
7.9.9 Cell lysis buffer for DNA extraction 225
7.9.10 ChIP-Lysis buffer 226
7.9.11 ChIP-Dilution buffer 226
7.9.12 Low salt buffer 227
7.9.13 High salt buffer 227
7.9.14 Lithium chloride buffer 227
7.9.15 TRIS-EDTA (TE) buffer 228
7.9.16 TRIS-Borate-EDTA (TBE) buffer (5x) 228
7.9.17 Elution buffer 229
7.9.18 Lysis buffer for 3C assay 229
7.9.19 Glucose-EDTA-TRIS (GET) solution 229
7.9.20 NaOH/SDS solution 230
7.9.21 Potassium acetate solution (3 M) 230
7.9.22 Z-buffer for β-galactosidase assay 230
8 Appendix 231
8.1 Conditions for ChIP-seq experiments 231
8.2 High stringency HIF-binding sites 231
8.2.1 High stringency HIF-1 binding sites in MCF-7 cells 231

8.2.2 High stringency HIF-2 binding sites in MCF-7 cells 239
8.2.3 High stringency HIF-2 binding sites in 786-O cells 246
8.3 DNA Sequence of “Nici” 256
8.4 Side by side alignment of the “Nici” sequence derived from clone (Clone IMAGE: 610879) with the reference genome HG18 257
8.5 ChIP-seq sequences of HIF-binding sites at 11q13.3 260
8.6 Abbreviations 261
8.7 Publications arising from this thesis 264
9 Bibliography 265
10 Declaration 274
11 Acknowledgments 275

1 CHAPTER 1: Introduction

1.1 The hypoxia inducible transcription factor (HIF) pathway

1.1.1 Atmospheric oxygen levels and evolution of life

After the formation of our planet the increase of atmospheric oxygen over millions of years due to photosynthesis evoked a variety of reactions from life on earth including mass extinction and adaptation. Initially molecular oxygen as a highly reactive compound oxidized inorganic minerals such as iron on the surface of the earth. Saturation of this buffer system over time led to a progressive increase in atmospheric oxygen levels over the last 3 billion years. The strong oxidative potential of oxygen proved toxic to most of uni-cellular life on earth (Dole 1965; Berner, Vandenbrooks et al. 2007; Semenza 2007). However, a newly developed form of life, the eukaryotes, which appeared around 1.6-2 billion years ago, managed to utilise molecular oxygen for energy production. Eukaryotes are characterised by well-defined intracellular compartments, such as the nucleus, which contains DNA. Furthermore, most of the eukaryotes contain mitochondria, developed from endosymbiotic prokaryotes which contain their own DNA (Margulis and Bermudes 1985; Cavalier-Smith 1992). Mitochondria produce ATP from ADP as an energy resource for the whole cell by a process called oxidative phosphorylation or aerobic respiration. This process involves energy production via electron transfer from glucose derivatives in the presence of oxygen. By using an oxidising and potentially toxic molecule this form of energy production is far more efficient than anaerobic glycolysis alone which is used by most prokaryotes. Therefore, aerobic respiration was and remains critical for the evolution of all higher forms of life on earth.

Importantly, over the last 0.5 billion years, the period of metazoan evolution, atmospheric oxygen levels varied greatly between 15 and 30% and influenced development of species (Berner 2009). In the light of these changes, species had to evolve

mechanisms to effectively use oxygen and to adapt to fluctuating levels in the short and the long term. For example, gigantism of some organisms could have been driven by the abundant supply of molecular oxygen around 300-260 million years ago. Indeed in recent experiments with fruit flies, hyperoxia (oxygen levels higher than the current 21%) over generations leads to an increased body size of the fly population. This observation could be explained by a hyperoxia driven selection of genes and the effect was linked to advantages in development and reproduction (Harrison and Haddad 2011).

During evolution higher organisms evolved in which oxygen can be transported through the body via circulating systems to every cell of the organism ensuring sufficient oxygen supply. In times of lower atmospheric oxygen evolution pressure clearly favoured organisms which possessed mechanisms of optimising oxygen and energy utilization.

1.1.2 The oxygen sensing pathway

When oxygen levels in individuals become low (e.g. in circumstances in which oxygen consumption is high such as growth and exercise or oxygen supply is low such as at high altitude) higher organisms are able to react with measures aimed at increasing oxygen delivery from the lungs to the peripheral tissues and at reducing oxygen consumption within cells. On the one hand, these reactions involve release of endocrine and paracrine factors driving enhanced capacity for oxygen transport (e.g. erythropoietin) and to facilitate blood supply to regions of low oxygen levels (e.g. VEGF, vascular endothelial growth factor). On the other hand, mechanisms within the cell are favoured which help to maintain energy production, despite reduced oxygen consumption, by increasing anaerobic glycolysis. This is achieved by induction of the expression of genes involved in the glycolytic pathway and necessary for the uptake of glucose as a substrate for anaerobic glycolysis.

The main driver of these transcriptional responses to low oxygen was identified as the hypoxia inducible transcription factor (HIF), which is stabilised under low oxygen tensions (Kaelin and Ratcliffe 2008; Semenza 2012). HIF is composed of a heterodimer of an α-subunit and a HIF-1β (ARNT, aryl hydrocarbon receptor nuclear translocator) subunit. The HIF-1β subunit is constitutively expressed, whilst the α-subunit (which can exist in several isoforms) is highly regulated at the level of protein stability (and in some organisms by transactivating activity). Oxygen-sensitivity is conferred by 2-oxoglutarate (2-OG) dependent dioxygenase mediated modification (hydroxylation) of residues on the HIF-α subunits, which either permit interaction with the von Hippel-Lindau E3 ubiquitin ligase (pVHL) and subsequent proteasomal degradation, or block interaction with the p300/CREB-binding protein (CBP) co-activator. When oxygen levels are low, hydroxylation is inhibited leading to high levels of HIF-α that dimerise with HIF-1β, bind to DNA and interacts with p300/CBP to activate the basal transcriptional machinery at target genes loci (Wang and Semenza 1993; Maxwell, Wiesener et al. 1999; Epstein, Gleadle et al. 2001; Ivan, Kondo et al. 2001; Jaakkola, Mole et al. 2001).

The HIFs, the prolyl hydroxylases and pVHL are all highly conserved in metazoans. For example, the system is conserved in Trichoplax adhaerens, the simplest animal (Loenarz, Coleman et al. 2011). During evolution various modifications were added to the system. For example, due to gene duplications vertebrates express at least three different prolyl hydroxylases (named after the homologue in C. elegans egl-9 (Egg laying defective family member nine) EGLN1, EGLN2 EGLN3 or PHD2, PHD1, PHD3) whereas lower organisms such as flies only have a single PHD isoform (egl-9) and amphibians express two isoforms. PHD2 shows the closest homology between the species and therefore is regarded as most related to the ancestral isoform (Rytkonen, Williams et al. 2011). It is also the most widely

expressed isoform in mammalian cells. HIF-α homologues are present in all metazoans. Vertebrates posses three isoforms: HIF-1α, HIF-2α and HIF-3α. Duplication of the ancestral *HIF* gene resulting in *HIF-1/2α* and *HIF-3α* genes occurred before the division of teleosts and mammals. pVHL and HIF-1β are also highly conserved among metazoans. Evidence for the crucial role of HIF in development is provided by the embryonic lethality induced by gene inactivation at each of the levels of controlling the HIF response (VHL, HIF-1α, HIF-1β (Maltepe, Schmidt et al. 1997), PHD2 (Takeda, Ho et al. 2006; Takeda, Cowan et al. 2007)) in mouse models.

During evolution another mode of controlling the HIF response appeared which consists of an asparagyl hydroxylase (*HIF1AN*, FIH-1, Factor inhibiting HIF-1α) that inhibits the transactivation potential of HIF by hydroxylating asparagine residues in the C-terminal transcativation domain (C-TAD or CAD) of HIF-α subunits, which interferes with the recruitment of cofactors (Lando, Peet et al. 2002). Its enzyme activity is also oxygen dependent but is more efficient under lower oxygen conditions than are the PHDs suggesting it functions as a gatekeeper for HIF activity. *HIF1AN* is found in mammals, but is absent in nematodes and flies. Interestingly, homologues of *HIF1AN* were found in beetles and more primitive organisms such as corals suggesting loss of this gene in some animals during evolution. This loss in nematodes and flies coincides with the loss of the C-TAD of HIF-α implying that regulation via FIH-1 was superfluous and therefore this gene was lost (Hampton-Smith and Peet 2009; Taylor and McElwain 2010).

In summary, the evolution of an oxygen sensing pathway in metazoans was necessary to adapt to atmospheric changes in oxygen tensions. In higher species gene duplications led to a more complex oxygen sensing system that allows them to react to imbalances in the sophisticated oxygen delivery system.

1.1.3 Oxygen levels and HIF in humans

Tissue oxygen tensions in multi-cellular organisms vary greatly and are dependent on oxygen supply and oxygen consumption in the specific tissues and cells. Although normal atmospheric oxygen is currently 21% and most scientific experiments in monolayer cell culture are carried out at this oxygen concentration this does not generally reflect oxygen tensions in tissues of complex animals. Tissue oxygen levels are highly dependent on the balance between oxygen transport from the lungs to the distal tissues and on oxygen consumption in those tissues. Efforts have been made to measure tissue oxygen tensions. For example, oxygen tensions in the kidney cortex were determined between 1.9 and 6 kPa depending on the method and the species examined (Dyson and Singer 2011) compared to 21 kPa for atmospheric pO_2. This suggests that even under normal atmospheric conditions ("normoxia") local oxygen tensions in the tissues are substantially lower than 21 kPa and changes in supply or demand could easily lead to a further reduction in oxygen ("hypoxia"). Indeed many pathological and patho-physiological conditions lead to acute or chronic lack of oxygen and therefore drive a hypoxic response via the HIF pathway.

Early observations in humans exposed to lower oxygen concentrations at high altitude suggested that hypoxia evokes responses in the body which lead to mechanisms that counteract the hypoxemia by increasing the production of red blood cells in the bone marrow and therefore the capacity of oxygen transport in the blood (Viault 1890; Miescher 1893). The humoral mediator underlying this reaction is erythropoietin (EPO) and analysis of messenger RNA and protein demonstrates that the predominant sites of production are in the kidney and the liver (Jelkmann and Bauer 1981; Tan, Eckardt et al. 1992). Analysis of the molecular mechanisms underlying this response revealed a

regulatory genomic region 3' from the *EPO* gene that conferred oxygen-dependent regulation. It was through the identification of the hypoxia-inducible transcription factor binding to this region that HIF (hypoxia-inducible factor) was discovered. The identified binding sequence in the genome was termed the hypoxia-response element (HRE).

1.1.4 PHD-VHL-HIF pathway in humans

In humans three HIF-α subunits exist. HIF-1α was first described in the context of hypoxic EPO regulation and was cloned in 1995 by G. Semenza's group and was shown to be member of the basic-helix-loop-helix-PAS (Per-ARNT-SIM) transcription factor family (Wang and Semenza 1995; McIntosh, Hogenesch et al. 2010). Subsequent analyses showed that HIF-1α was widely expressed in virtually all cell-types, irrespective of their ability to express EPO, indicating additional functions for this oxygen-sensing apparatus and additional target genes such as *PGK-1* (phosphoglycerate kinase 1) and *LDHA* (lactate dehydrogenase A) were soon identified (Firth, Ebert et al. 1994; Semenza, Roth et al. 1994).

The existence of HIF-2α was described in 1997 by different groups (Ema, Taya et al. 1997; Flamme, Frohlich et al. 1997; Tian, McKnight et al. 1997) and its distribution appears to be more restricted to certain cell types, such as endothelial cells or interstitial cells in the kidney cortex (Wiesener, Jurgensen et al. 2003). A third isoform, HIF-3α, has been described in 1998 (Gu, Moran et al. 1998). It is inducible under hypoxic conditions but appears to undergo intensive splicing, and its functional role in the hypoxic response is poorly understood (Maynard, Qi et al. 2003; Tanaka, Wiesener et al. 2009). Recent reports point to a dominant-negative function for HIF-3α splice variants with respect to the other HIF-α isoforms (Maynard, Evans et al. 2005).

All HIF-α subunits bind HIF-1β via protein-protein interaction at the PAS domain (Figure 1.1). HIF-1β is a constitutively expressed protein, which can also interact with other binding partners such as the aryl hydrocarbon receptor (AhR) (McIntosh, Hogenesch et al. 2010). The heterodimers of HIF-1 and HIF-2 are able to interact with DNA at specific binding sites. Oxygen sensing in the HIF-α molecules occurs C-terminal to the PAS domain and is mediated via hydroxylation of two prolyl residues (Pro 402 and Pro 564 for human HIF-1α and an asparagine residue (As 803)). The prolyl residues are located within an oxygen-dependent degradation domain (ODDD) with N- and C-terminal portions (NODDD, CODDD).

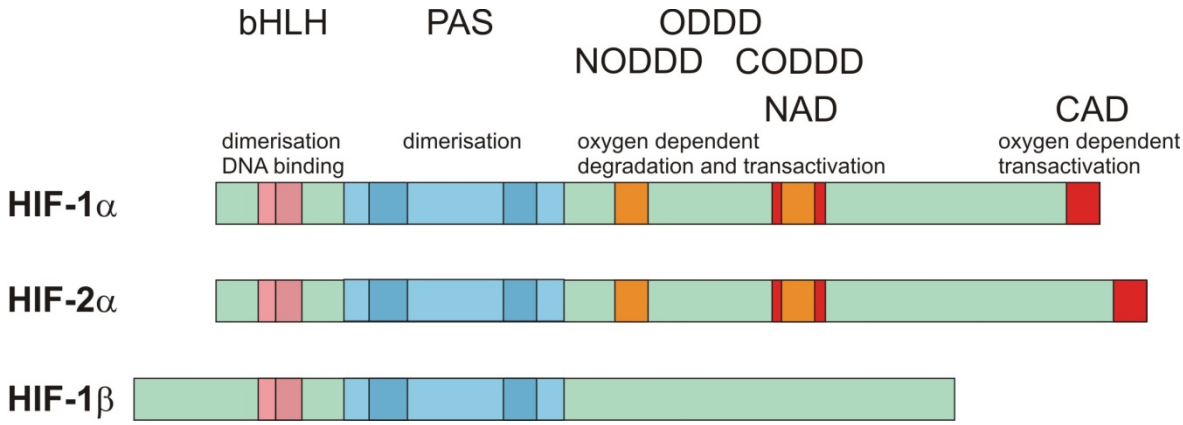


Figure 1.1: Schematic structure of HIF proteins. bHLH: basic-helix-loop-helix, PAS: per-arnt-sim, ODDD: oxygen dependent degradation domain, NODDD: n-terminal ODDD, CODDD: c-terminal ODDD, NAD: n-terminal activation domain, CAD: c-terminal activation domain.

Hydroxylation of HIF-α subunits is the rate-limiting step in their normoxic degradation. Once hydroxylated, HIF interacts with the pVHL-E3 ubiquitin ligase in a manner that is highly specific compared to HIF containing non-hydroxylated prolyl residues (Hon, Wilson et al. 2002). The pVHL ubiquitin ligase complex then ubiquitinates HIF, in a process that targets it for rapid proteasomal degradation (Maxwell, Wiesener et al. 1999).

Hydroxylation at each site may be catalysed by one of three prolyl hydroxylases (PHD1-3) which are dependent on molecular oxygen, 2-oxoglutarate, Fe^{2+} and ascorbate (Nytko, Spielmann et al. 2007). Expression analysis and enzyme activity assays point to a crucial role for PHD2 in the hydroxylation process. Indeed, PHD2 is widely expressed whilst PHD1 and PHD3 have a more restricted tissue expression pattern. Also, PHD2 knockout is embryonic lethal in mice underscoring the importance of this enzyme (Takeda, Ho et al. 2006).

It is of note that both PHD2 and PHD3 are themselves transcriptional targets of the HIFs and so constitute a negative feedback loop for the HIF response (Epstein, Gleadle et al. 2001; del Peso, Castellanos et al. 2003). Increased expression of PHD2 and PHD3 leads to an increased enzyme activity in low oxygen concentrations, which limits or acclimatises HIF activity in prolonged hypoxia.

1.1.5 Effects of HIF on gene expression

The oxygen sensing mechanism enables cells and the whole individual to adapt to non-favourable conditions. This is important to secure cell survival or to induce controlled cell death if necessary. In order to initiate these responses HIF-dimers bind to designated HREs in the genome and together with cofactors stimulate the transcriptional machinery to modulate gene expression. In addition to the examples given above, a multitude of HIF-target genes have now been identified. These include secreted factors such as EPO, VEGF and PDGF to increase oxygen delivery to the cells and transporters and glycolytic enzymes to maintain energy balance in conditions when oxygen levels are low. Other factors important for metal homoeostasis, cell cycle control and lipid metabolism are also under the control of HIF transcription factors.

Pan-genomic analyses of the hypoxic response using gene expression arrays in different cell lines have reported that several hundred up to thousands of genes are either up or downregulated in hypoxia (or by hypoxia mimetics, which chemically activate HIF by inhibiting PHD activity). However, the existence of multiple HIF-α isoforms with distinct expression patterns raises fundamental questions as to what extent each isoform is contributing to the hypoxic response and whether there are differences in the transcriptional pathway induced by the different isoforms.

Gene knockout studies in mice confirmed the importance of the HIF system for the mammalian development. HIF-1α knockout leads to mid-gestational embryonic lethality and HIF-2α knockout also causes embryonic lethality, blood vessel defects and insufficient lung maturation (Iyer, Kotch et al. 1998; Ryan, Lo et al. 1998; Tian, Hammer et al. 1998; Peng, Zhang et al. 2000; Compernolle, Brusselmans et al. 2002). The phenotype of HIF-2α knockout mice varies widely depending on the background with some individuals even surviving to term (Compernolle, Brusselmans et al. 2002; Scortegagna, Ding et al. 2003). Although there could be a substantial redundancy in target genes due to high sequence homology of HIF-1α and HIF-2α and binding to the same dimerisation partner, these findings indicate that each isoform is necessary for the successful development of mammals.

Therefore, HIF knockdown studies in different human cell lines or experiments in murine embryonic fibroblast cells deficient for single HIF isoforms have been performed to elucidate the target gene profile of each isoform. The broad tissue distribution of HIF-1α and early studies on single genes had indicated that this is the main mediator of hypoxic gene regulation and the main driver of hypoxic induction of glycolytic genes. These findings were largely confirmed in comprehensive studies investigating the normoxic and

hypoxic transcriptomes of HIF-1α deficient or competent cells (Covello, Kehler et al. 2006; Elvidge, Glenny et al. 2006; Warnecke, Weidemann et al. 2008).

However, early analysis of HIF targets had also indicated that the HIF-2α isoform also has transcriptional activity at least at selected gene loci such as *EPO* and *VEGFA*. Subsequent DNA microarray studies confirmed that HIF-2α has transcriptional activity on a broad range of target genes at least in a context in which it is the only expressed isoform (Hu, Wang et al. 2003; Wykoff, Sotiriou et al. 2004). However, HIF-1α is expressed in nearly all cells examined so far and in the context of HIF-1α competent cells, experiments manipulating the HIF-2α pathway by knockdown or knockout result in a very restricted number of genes responding to HIF-2α (Elvidge, Glenny et al. 2006; Hu, Iyer et al. 2006; Warnecke, Weidemann et al. 2008). Genes exclusively responding to HIF-2α are rare. Interestingly, these include *EPO* in Hep3B and Kelly cells (Warnecke, Zaborowska et al. 2004) or *Oct-4*, a transcription factor important to maintain stem cell pluripotency (Covello, Kehler et al. 2006). Taken together, in terms of gene transcription HIF-1α plays the major role in hypoxic gene regulation whilst HIF-2α selectively targets genes depending on the cellular context.

These findings are consistent with an evolutionary process in which the HIF-isoforms arise from gene duplication of a precedent HIF gene. HIF-1α is widely expressed and is conserved to the earliest appearance of HIF in single cell organisms and is therefore likely necessary for individual cell survival. HIF-2α and more so HIF-3α (as far as known) have a very restricted expression pattern in mammals and their function is limited compared to HIF-1α. Nevertheless, knockout of HIF-2α is still lethal in mice (HIF-3α knockout mice have not been reported) and the regulation of EPO and VEGF as humoral factors would

suggest that HIF-2α has taken over novel, more specialised functions mandated by the acquisition of complex body morphologies. In respect to gene evolution, these findings would categorise HIF-2α and HIF-3α as sub functionalised or neofunctionalised (Nowick and Stubbs 2010).

1.2 HIF in cancer

The physiological role of HIF is to maintain oxygen and energy homoeostasis and to secure successful growth and development. One important pathological example of unrestricted growth and oxygen demand is cancer. As the tumour grows it outstrips its blood supply leading to intratumoral hypoxia and activation of HIF pathways. Sufficient neo-formation of blood vessels (angiogenesis) to guarantee supply of oxygen and nutrients even to the most peripheral regions of fast growing tumours is one paradigm of HIF function in cancer. HIF induces the expression of growth factors such as VEGFA or PDGF to guide vessel formation to these regions (Forsythe, Jiang et al. 1996; Schito, Rey et al. 2012). In addition, HIF action can promote tumour progression by inducing genes shifting tumour metabolism towards anaerobic glycolysis which leads to a stable energy balance at reduced oxygen consumption and to shunting of metabolites into pathways necessary for cell proliferation. Evidence for the activity of these mechanisms in cancer comes from a multitude of studies showing that HIF is indeed overexpressed in tumours and that its expression in most cases is associated with a poor outcome. These cancers include tumours of the gastrointestinal system, the brain, the lung, the skin, the prostate and the kidney. Interestingly, most tumours express both HIF-α isoforms, whereas in non-tumour tissue HIF-2α expression is restricted to a smaller number of cell types than HIF-1α (Wiesener, Turley et al. 1998; Wiesener, Jurgensen et al. 2003).

1.3 VHL syndrome and renal cancer

Since rapidly growing tumours can become hypoxic leading to high levels of HIF, cause and effect can be difficult to distinguish in the association between HIF level and tumourigenicity. One important cancer type that provides a paradigm for studying the HIF pathway dissociated from tumour hypoxia is VHL-associated clear cell renal cancer - loss or inactivation of VHL leading to constitutive activation of HIF even in well oxygenated regions of the tumour.

VHL-syndrome itself is an autosomal dominant inherited, monogenic disease in which a germline mutation in one *VHL* allele and a subsequently acquired mutation in the remaining allele lead to the complete loss of function of pVHL resulting in abnormal high levels of HIF. The syndrome comprises renal cell carcinoma (RCC), highly vascularised tumours (hemangioblastoma) in the retina, cerebellum or spinal cord and pheochromocytoma. VHL disease can be sub-classified according to phenotype (Table 1.1) and genotype-phenotype correlations exist. Subtype 1 comprises risk for hemangioblastoma and high risk for RCC with no increased risk for pheochromocytoma. Subtype 2 includes increased risk for hemangioblastoma and pheochromocytoma together with no increased risk for RCC (Type 2A) or high increased risk for RCC (Type 2B). Type 2C shows high risk for pheochromocytoma and no increased risk for RCC or hemangioblastoma. Of note, deletions or mutations of *VHL* affecting the interaction with HIF-α and leading to HIF stabilisation are strongly associated with VHL disease types 1 and 2B which are characterised by high risk for RCC (Knauth, Bex et al. 2006; Li, Zhang et al. 2007).

VHL subtype	VHL mutation	Hemangioblastoma	Renal cell carcinoma	Pheochromozytoma	HIF levels
1	Deletion, nonsense, missense	+	+	-	+++
2B	Missense	+	+	+	++
2A	Missense	+	-	+	+
2C	Missense	-	-	+	-

Table 1.1: Subtypes of the VHL-syndrome (modified from (Kaelin 2007)).

The existence of mutations which do not compromise HIF degradation but lead to a phenotype in VHL disease points to additional functions of pVHL outside the regulation of HIF. Indeed, additional functions of pVHL have been described in the regulation and stabilisation of the primary cilium (Wiesener, Maxwell et al. 2009). In tubular cells of the kidney VHL loss of function can lead to the loss of the primary cilium which is crucially involved maintaining cell polarity. The disturbance of which can lead to renal cyst development, a precancerous condition. Other non HIF-dependent functions of pVHL involve regulation of microtubules, extracellular matrix and potentially regulation of p53 (Frew and Krek 2007).

Acquired *VHL* mutations or hypermethylation of the *VHL* promoter that lead to inhibition of VHL mRNA expression can be detected in 50-80% of sporadic cases of renal carcinoma. The most prominent effect of VHL inactivation is the disruption of the degradation pathway of HIF. HIF-α isoforms are stabilised in a normoxic environment and together with HIF-1β start a pseudo-hypoxic transcriptional response. This involves high production of VEGFA which leads to increased vessel formation in part explaining the high vascularisation of pVHL-defective tumours. The VEGFA pathway is an established therapeutic target in renal cancer. Furthermore, increased glucose uptake via expression of glucose transporter 1 provides metabolites for energy production and cell growth.

Genes involved in glycogen and lipid metabolism (Gimm, Wiese et al. 2010; Pescador, Villar et al. 2010) lead to increased production, uptake and storage of these molecules in the cells, contributing to the characteristic "clear-cell" phenotype of most renal cell carcinomas (ccRCC) in histopathological examinations.

1.4 The divergent roles of HIF-1 and HIF-2 in renal cancer

The role of VHL, HIF-1 and HIF-2 in tumour formation and progression has been examined in several studies. In VHL disease kidneys contain many precancerous nests of cells that have lost pVHL and consequently have high levels of HIF (Mandriota, Turner et al. 2002). Few of those lesions progress to an advanced cancer. In cell culture experiments acute loss of VHL leads to cellular senescence (Young, Schlisio et al. 2008). Both findings indicate that additional mechanisms are necessary to initiate proliferation and growth. Immunohistochemistry experiments showed that HIF-1α expression is frequently lost in VHL-defective cancer, but HIF-2α expression is always detectable (Gordan, Lal et al. 2008). This finding is in contradistinction to the normal expression pattern of HIF in the kidney, in which HIF-1α can be detected in tubular cells, whereas HIF-2α is only detectable in interstitial fibroblasts and glomerular endothelial cells (Rosenberger, Mandriota et al. 2002). This finding suggests that VHL loss releases HIF-2α expression in the tumour cells and that HIF-2α contributes to tumour formation and progression.

Tumour xenograft experiments in mice showed that reintroduction of functional VHL in renal cancer cells reduces tumour burden. However, forced overexpression of HIF-2α is able to override this effect and restores tumour growth. HIF-2α knockdown in VHL-defective cells is sufficient to reduce tumour size (Kondo, Klco et al. 2002; Kondo, Kim et al. 2003). In another study, overexpression of either HIF-1α or HIF-2α in the context of

VHL-deficiency was performed. Interestingly, HIF-1α slowed down tumour growth, whereas HIF-2α led to a further increase in tumour size (Raval, Lau et al. 2005). Furthermore, in a series of experiments Kaelin and co-workers examined the functional effects of HIF-1α in renal cancer (Shen, Beroukhim et al. 2011). They showed that in addition to deletions on chromosome 3p (chromosomal location of *VHL*) deletions of 14q, the chromosomal region, which encompasses *HIF1A*, were more common in renal cancer compared to other types of cancers (approximately 45%). In addition to deletions, the detection of HIF-1α mRNA variants in some renal cancer cell lines indicates re-arrangements at the *HIF1A* locus with potential loss of function. To further asses the functional relevance of HIF-1α the authors re- or overexpressed HIF-1α in different renal cancer cell lines and found growth inhibiting effects in cells which previously had lost HIF-1α expression (Shen, Beroukhim et al. 2011). In wild-type HIF-1α competent cells no effect was seen. When knocking down HIF-1α in these cells, increased tumour progression was observed. The authors further showed that HIF-1α deletion mutants occurring in renal cancer have weaker effects on tumour growth inhibition than wild-type HIF-1α. Taken together, these findings indicate a role for HIF-1α as a potential tumour suppressor and a role for HIF-2α as an oncogene in VHL-defective renal cancer.

1.5 Mechanisms of differential functions of HIF-1α and HIF-2α

The exact mechanisms by which each HIF isoform exerts its specific role in renal cancer are not entirely clear. Opposing effects of each isoform on the activity of other oncogenic factors such as p53 or mTOR (mammalian target of rapamycin) have been described. In addition, in renal cancer a role for HIF-2α in promoting c-MYC (avian myelocytomatosis viral oncogene homolog) activity has been demonstrated, whereas HIF-1α appeared to reduce

c-MYC activity (Gordan, Bertout et al. 2007). c-MYC is a basic helix-loop-helix/leucine zipper transcription factor involved in cell-cycle regulation and cell growth and is frequently overexpressed in various tumours. The mechanism by which the HIF-α subunits exert their effects was shown to be independent from their transcriptional activity. Gordan and colleagues observed a differential interaction of HIF-α subunits with binding partners of MYC, namely SP1 (Specificity protein 1) and MAX (MYC-associated factor X) (Gordan, Bertout et al. 2007). HIF-1α interacted with both whereas HIF-2α only bound to MAX. The authors concluded that the HIF-1α/MAX complex inhibited transcriptional activity of MYC/MAX dimers leading to downregulation of cell-cycle related genes such as *p21* and that HIF-2α bound to MAX increased expression of these genes. These insights indicate a role for HIF-α subunits in addition to direct transcriptional activation.

However, given the fact that the main HIF-response is characterised by the transcriptional programme directly evoked by HIF-binding it is likely that differential expression profiles also contribute to the different functions of HIF-1α and HIF-2α. Single gene and expression array analyses have identified a number of genes, which are regulated via both or one or the other HIF isoform. In renal cancer for example, Glut1 and VEGFA are regulated by both HIF-α isoforms, whereas glycolytic enzymes such as HK1, ALDOA or PGK1 are mainly regulated via HIF-1α. TGFA and cyclin D1 are exclusively targets of HIF-2α in renal cancer (Hu, Wang et al. 2003; Wykoff, Sotiriou et al. 2004; Raval, Lau et al. 2005).

The mechanisms underlying this target gene specificity remain uncertain. One possibility is that HIF-1α and HIF-2α bind to different sites in the genome. The core-binding motif has been identified as RCGTG (R = A or G) (Wenger, Stiehl et al. 2005) based on a limited

number of published HIF-binding sites for both HIF-isoforms across a range of species. The fact that HIF-1α and HIF-2α can regulate the same genes in the same cell line suggests that both bind to a similar recognition site in the respective regulatory element of this gene. Indeed, recent reports indicate that HIF-1α and HIF-2α regulate the expression of PHD3 in renal cancer and other cells via binding to the same intronic HRE (Lau, Tian et al. 2007). In the same report another HIF target gene, *CA9*, was examined which has been shown to be regulated exclusively via the HIF-1α pathway. Interestingly, it could be shown that in renal cancer cells HIF-2α is able to bind to the regulatory sequence in the *CA9* promoter, but has no effects on gene transcription (Lau, Tian et al. 2007) suggesting post-binding mechanisms of transcriptional selectivity. Using chimeric HIF proteins, the authors determined that the C-terminal protein sequence of HIF-α was crucial for induction of CA9 expression. In a similar study Hu and colleagues also found that both HIF-α isoforms can bind to the same DNA sequence, but still have distinct activity on target gene expression (Hu, Sataur et al. 2007). In overexpression studies of hybrid proteins they identified that the N-TAD was responsible for conferring target gene selectivity. Overall, these studies suggest that it is not the ability of DNA-binding, but post DNA-binding mechanisms which define the transcriptional output of the different HIF isoforms.

Another important finding with respect to differential HIF-α transcriptional activity comes from studies on the C-TAD. In particular, FIH-1 (*HIF1AN*) has been shown to be more active on the C-TAD of HIF-1α than on the C-TAD of HIF-2α due to a single conserved amino acid substitution within the CAD (Bracken, Fedele et al. 2006). Therefore, FIH-1 mediated inhibition of p300 recruitment and transactivating activity is more efficient on HIF-1α than HIF-2α.

A recent report supports the hypothesis that FIH-1 is involved in the different transcriptional responses to HIF-1α and HIF-2α in renal cancer (Khan, Bhattacharyya et al. 2011). The authors showed that inhibition of FIH-1 activity induced HIF-1α driven gene expression leading to increased apoptosis in renal cancer cells. No effects were observed in a cell line exclusively expressing HIF-2α. This suggests that FIH-1 mediated inhibition of HIF-1α is partly responsible for repressing the anti tumourigenic effects of HIF-1α in renal cancer and further supports the hypothesis that additional factors beyond HIF-binding substantially contribute to target gene selectivity.

1.6 Genetic alterations in renal cancer

1.6.1 Rare familial syndromes

Whilst rare monogenic tumour syndromes have shed light on the molecular mechanisms underlying many common cancers including RCC, insights arising from large-scale genome-wide association studies have to date been less informative. For example, the observation that VHL-deletion results in upregulation of HIF under normoxic conditions led to the description of the interaction between pVHL and HIF in order to target HIF for proteasomal degradation (Maxwell, Wiesener et al. 1999). The finding that HIF-upregulation is associated with renal tumour formation and progression suggests that HIF plays a crucial role in renal tumour biology. In fact, over 80% of renal cancers accounting for more than 100,000 deaths per year worldwide are associated with the loss of VHL function and HIF-activation (Ferlay, Shin et al. 2010).

Other malignant tumours of the kidney deriving from tubular cells are also associated with genetic alteration of tumour suppressor genes and oncogenes. Type 1 papillary renal cell carcinoma are characterised by a gain-of-function mutation in the *MET* gene which

codes for a transmembrane receptor tyrosine kinase. Constant activation of this pathway triggers tumourigenic processes that include cell migration and mitosis. Hereditary type 2 papillary renal cell carcinoma and renal cysts are associated with the loss of fumarate hydratase (FH), an enzyme involved in the Krebs-cycle (Tomlinson, Alam et al. 2002). The exact mechanism how loss of FH leads to tumours is not fully understood, although a clear link to the HIF system exists. FH deficiency leads to accumulation of fumarate which can act as an inhibitor of the prolyl-hydroxylases via competition with the 2-oxoglutarate cofactor (Isaacs, Jung et al. 2005). HIF-α stabilisation is frequently observed in FH deficient tumours. However, in animal models of FH deficiency development of cysts was not reduced by knockout of HIF-α isoforms, suggesting that other mechanism are involved in cyst development and progression (Adam, Hatipoglu et al. 2011). However, in common with VHL-deficient mice, FH-deficient mice do not develop renal tumours. Therefore, so far the role of HIF in VHL- or FH-associated tumour development or progression cannot be examined in animal knockout models.

1.6.2 Genome-wide approaches in sporadic RCC

More recently, genome-wide attempts to better understand the basic mechanisms of renal cancer have identified a variety of other candidate genes potentially involved in the pathogenesis of RCC. In a PCR-based sequencing study of histone modifying enzymes performed on tumour samples of multiple cancer types Van Haaften et al identified inactivating mutations in the coding region of the *UTX* gene, a histone 3 lysine 27 demethylase (van Haaften, Dalgliesh et al. 2009). This is one of a large number of enzymes that modify the tails of histone proteins and control the structure and function

of chromatin, potentially playing a role in gene regulation. Mutations were discovered in a large variety of cancer types that included renal cancers.

Subsequent results from the same group provide further evidence of a role for histone modifiers in the development of renal cancer (Dalgliesh, Furge et al. 2010). Sequencing of the coding regions of 3544 genes in approximately 100 clear cell renal tumours and follow-up sequencing of candidate genes in another 311 tumours revealed that two additional histone modifying enzymes, SETD2, a histone 3 lysine 36 methyltransferase, and JARID1C a histone 3 lysine 4 demethylase had truncating mutations in 3% of the cancers tested. In this study VHL mutations were found in 55% of samples which is comparable to previously reported rates. All mutations in the new candidate genes were associated with VHL-deficiency. Subsequently, exome wide sequencing identified *PBRM1* a member of the SWI/SNF chromatin remodelling complex as frequently (35%) mutated in ccRCC (Varela, Tarpey et al. 2011). The results for mutations in chromatin modifiers or associated proteins have been confirmed in an independent study in which additional low frequency mutations in other members of the ubiquitin-mediated proteolysis pathway were also observed (Guo, Gui et al. 2012). In summary, these results indicate that besides the well-established role of VHL mutations in renal cancer, other mutations especially in the chromatin modifier pathway may play important roles in ccRCC development. However, individually each gene is mutated in only a small proportion of renal tumours. In addition, these approaches only studied 3000 out of approximately 20000 genes and were restricted to identifying mutations in the coding sequences.

Since there is a high overlap between mutations in these newly identified genes and VHL loss it is therefore conceivable that the deregulation of these genes could modulate the HIF pathway. Importantly, HIF-activation can be detected in the earliest pre-cancerous

lesions in VHL-patients indicating that loss of VHL and consequent HIF activation is the first genetic event. Together with the finding in cell culture experiments that acute loss of VHL leads to cell senescence a model would be conceivable in which other genetic alterations are necessary to overcome the senescence phenotype and to "start" tumour progression.

A more conventional approach to identifying genetic diversity that is linked to disease is to examine for an association with germline variants at single nucleotide polymorphisms (SNPs). Importantly, a recent genome-wide association study (GWAS) testing for the susceptibility to renal cancer identified predisposing SNPs within the first intron of the *EPAS1* gene (encoding for HIF-2α) (Purdue, Johansson et al. 2011). These SNPs are in low linkage disequilibrium with SNPs in the coding regions of the *EPAS1* gene and so their functional consequence is not yet clear. However, the association between RCC associated polymorphisms and a major transcriptional driver in ccRCC is striking and further emphasises the importance of the HIF pathway and HIF-2α in particular in the pathogenesis of RCC.

Importantly, in the same GWAS analysis another SNP on chromosome 11, remote from any annotated genes, but covered by the SNP arrays, reached high genome-wide significance. This SNP has a minor allele frequency of about 7% and is protective against renal cancer. Prior to my work, the significance of this site was unknown. However, since both this region and polymorphisms at the *EPAS1* locus are not linked to coding polymorphisms they imply an effect on gene regulation rather than protein function and further suggest that approaches that examine only coding regions of the genome will likely miss important effects.

1.7 Profiling of the hypoxic response and defining HIF-binding sites

Given the importance of the HIF response to cellular biology in general and the pathogenesis of cancer (especially renal cancer) in particular, many attempts have been made to determine the transcriptional targets of HIF in a variety of cellular settings. However, genome-wide approaches incorporating expression microarrays will not distinguish direct transcriptional targets from indirectly regulated genes, whilst studies at individual gene loci have been limited to just a few genes. For example, direct transcriptional regulation of the prototypical *EPO* gene is effected through binding of HIF to a hypoxia response element (HRE) that lies in the 3' UTR of the gene. Furthermore, binding of HIF within the first intron of *EGLN3/PHD3* affects transcriptional control at this locus. The existence of remote regulatory regions in the genome is well documented for regulatory elements 40-50 kb upstream of the *α-globin* locus (Higgs, Engel et al. 2012). These regions are influencing the expression of α-globin and mutations in these loci have been described that lead to α-thalassaemia.

The identification of the HIF-binding site remote from the promoter highlights the difficulty in predicting HIF-binding bioinformatically. Being short - just 5 base pairs long - the HIF-binding HRE motif is abundant across the human genome, with in excess of 1.1 million occurrences. Since HIF can bind and act through HREs remote from the promoter, potential binding sites are present at virtually every gene and therefore do not predict direct from indirect transcriptional targets of HIF. Furthermore, at a given gene locus, the presence of multiple potential binding motifs can make identification of the actual binding site difficult, even experimentally by conventional candidate approaches.

Recently, technological advances in the analysis of protein-DNA interactions and DNA sequences have allowed the screening of whole genomes for transcription factor binding

sites irrespective of their distance from promoters. For example, the proliferation and function of breast cancer cells is highly dependent on the activity of estrogen receptors, which are targeted in breast cancer therapy. However, the full direct transcriptional response to estrogen receptor binding remained obscure until recently. In 2005 a study using the combination of chromatin immunoprecipitation, a method which allows analysis of protein-DNA interactions, and tiled DNA microarrays covering the whole DNA sequences of chromosomes 21 and 22 identified approximately 50 binding sites across these chromosomes in MCF-7 breast cancer cells (Carroll, Liu et al. 2005). Importantly, the authors also detected frequent estrogen receptor binding to loci remote from annotated genes which were still associated with estrogen mediated regulation of genes in the vicinity. This finding suggested the common existence of transcriptional enhancers in the estrogen receptor binding sites repertoire. Only two years later the same authors presented approximately 3500 estrogen binding sites in MCF-7 breast cancer cells on a whole genome level using tiled microarrays (Carroll, Meyer et al. 2006). More recently, chromatin immunoprecipitaion coupled to high-throughput DNA sequencing (ChIP-seq) has allowed an unbiased approach to identify estrogen binding sites which is not restricted to the coverage provided by a tiled microarray (Fullwood, Liu et al. 2009). Furthermore, sequencing of DNA bound to epigenetic (i.e. not encoded in the sequence of the DNA) markers such as histone modifications and RNA polymerase 2 or analysis of long-range interactions of the estrogen receptor allowed a comprehensive description of the estrogen receptor interactome in breast cancer cells. This rapid development in the genome-wide analysis of the estrogen receptor exemplifies how modern technology helps to understand the genomic and epigenomic landscape of transcription factor binding.

One major finding of these studies was that transcription factor binding is not restricted to promoter regions of genes. Moreover, estrogen receptor binding to remote regions is observed much more frequently than binding to promoter regions (only 4% of estrogen receptor binding sites were detected within 1 kb of promoters of annotated genes) (Carroll, Meyer et al. 2006). These findings imply that in general transcription factor binding to non promoter regions is frequent and functional. Furthermore, for the HIF system the *EPO* enhancer at the 3' end of the gene and the intronic enhancer in the *EGLN3* gene (Pescador, Cuevas et al. 2005; Lau, Tian et al. 2007) presents proof-of-principle for a long-range mechanism of hypoxic gene regulation.

True pan-genomic analyses of protein binding and histone modification, effected by ChIP-seq methodologies, also permit both quantitative and spatial description of these features at a typical or "average" gene locus. Multiple datasets can then be correlated to identify patterns associated with factors such as gene expression and/or transcription factor binding.

In a thorough and comprehensive analysis of epigenetic modifications by ChIP-seq coupled with expression profiling in CD4+ lymphocytes, Barski et al showed that distinct epigenetic profiles are present at gene loci. For example, histone 3 lysine 4 trimethylation (H3K4me3) is associated with promoters of actively transcribed genes, whereas histone 3 lysine 9 trimethylation (H3K9me3) at promoters correlates well with repressed genes (Barski, Cuddapah et al. 2007). Similarly, high levels of histone 3 lysine 36 trimethylation (H3K36me3) across the gene body are found at active genes, whereas high levels of histone 3 lysine 27 trimethylation (H3K27me3) are found at repressed genes. However, the extent to which these modifications either influence or merely reflect gene transcription is not fully understood.

The epigenetic findings were also expanded to characterise regulatory elements remote from gene loci, for which an additional distinct pattern of histone modification was found (e.g. high levels of histone 3 lysine 4 monomethylation (H3K4me1) and histone 3 lysine 27 acetylation (H3K27ac) with low levels of the promoter mark H3K4me3) (Creyghton, Cheng et al. 2010; Rada-Iglesias, Bajpai et al. 2011). In addition, the presence of RNA polymerase 2, the main component of the basal transcriptional machinery and the presence of the transcriptional co activator p300/CPB, which in the HIF-system is recruited via the C-TAD also indicated active enhancers (Heintzman, Stuart et al. 2007). Interestingly, the distribution of active enhancers across the genome is highly cell type specific, whereas the activity of promoters is more uniform between cell lines (Heintzman, Hon et al. 2009). This suggests that enhancers more than promoters are able to shape the specific transcriptional profiles of different cell types.

To date there has been relatively sparse pan-genomic analysis of the dynamic changes in transcriptional rate and histone modification in response to transcription factor binding and transcriptional activation. The large dynamic range in HIF levels and the rapid response of HIF pathways to hypoxia make it an ideal model system in which to study such changes. In addition, a number of histone modifying enzymes are either direct transcriptional targets of HIF and/or require oxygen as a co substrate furthering the potential for hypoxic regulation of the epigenetic environment and increasing the utility of such analyses.

The technical advances in genome-wide studies together with advances in computational management of these data sets now allow a comprehensive assessment of transcription factor binding, epigenetic phenomenon and chromatin interactions.

1.8 Aim of this study

The importance of the HIF system for physiologic and patho-physiologic responses to hypoxia is well documented. In this thesis I aim to define HIF-binding sites across the genome using ChIP-seq and to match these sites with other newly generated or existing genome-wide data for genetic and epigenetic features of gene regulation. The new approach by ChIP-seq provides whole genome coverage, is more cost effective compared to tiled microarrays of the whole genome and has been shown to provide better resolution of the region of transcription factor binding sites (Park 2009).

The previously published expression analysis examining the HIF dependency of gene expression under hypoxia in MCF-7 cells and ChIP on chip analysis for HIF-binding in the same cell line provide high quality data sets with which to compare the results of the ChIP-seq method and to define valid thresholds. Therefore, the first part of the thesis (Chapters 2 and 3) describes analysis carried out in MCF-7 cells including ChIP(-seq) experiments against HIF-1β to further validate true HIF-α binding sites. Furthermore, additional ChIP-seq experiments for RNA pol2 and histone modifications as well as sequencing of the whole transcriptome (RNA-seq) were performed in normoxia and hypoxia to bring HIF-binding sites into their epigenetic and functional context.

The evolving role of HIF-2α in renal cancer development and progression prompted me to further analyse HIF-2 binding sites in renal cancer cell lines. This will be the subject of Chapter 4 which aims to increase the understanding of the target gene profile of HIF-2α in these tumours and to screen for potential candidate genes. ChIP-seq experiments for HIF-binding were performed using the renal cancer cell line 786-O. This cell line has lost functional VHL and only expresses the HIF-2α exhibiting common features of renal cell carcinoma cell lines. In Chapter 5, I present work on one specific HIF-binding locus (at

11q13.3) identified in 786-O cells. HIF-binding at this locus was unique to pVHL-defective RCC cells, effected a long-range interaction that regulated the *CCND1* promoter and was disrupted by a RCC protective haplotype spanning this region that was identified in a large-scale GWAS of RCC susceptibility.

These experiments will allow

a) the identification of HIF-binding characteristics within a cell line on a genome-wide level,
b) the comparison of binding properties of different HIF-α isoforms,
c) the comparison of HIF-binding between cell types,
d) the correlation with expression profiles in the same cell line,
e) the identification of important pathways directly affected by the HIF response in cancer cells,
f) the investigation of epigenetic characteristics at HIF-binding sites,
g) the identification of candidate HIF-binding sites and target genes relevant for tumour biology.

2 CHAPTER 2: Genome-wide HIF-binding in MCF-7 cells using ChIP-seq

2.1 Introduction

Before the use of genome-wide techniques most transcription factor binding sites have been discovered by analysis of single gene loci restricted mainly to intragenic or juxtagenic regions. For example, transgenic experiments examining the hypoxic regulation of erythropoietin led to the discovery of a 3' hypoxia response element (HRE), which facilitated the identification of hypoxia-inducible factor (HIF) itself (Wang and Semenza 1995). Likewise, a bioinformatic approach uncovered the existence of a functional HRE in the first intron of the *EGLN3* gene many thousand kilo bases downstream of the promoter which is responsible for hypoxic induction of expression (Pescador, Cuevas et al. 2005). However, more recently novel techniques such as ChIP-on-chip or ChIP-seq analysis have identified a myriad of transcription factor binding sites at loci far distant from promoters and genes (Carroll, Meyer et al. 2006; Heintzman, Stuart et al. 2007). Some of these remote loci may regulate the expression of neighbouring (or even non-neighbouring genes) over long genomic intervals (Fullwood, Liu et al. 2009; Schoenfelder, Clay et al. 2010). Although the mode of action of transcription factors bound to these enhancers is not entirely clear, a direct interaction of the transcription factor with the promoter of the regulated gene seems to be a prerequisite for transcriptional regulation. Furthermore, models in which promoters and enhancers are concentrated in so called transcriptional factories characterised by accumulation of RNA polymerase 2 have been suggested (Schoenfelder, Clay et al. 2010).

Microarray analyses using hypoxic or chemical induction of the HIF response in combination with siRNA mediated knockdown of HIF-1α or HIF-2α have defined the extent of HIF-dependent versus HIF-independent regulation of transcript abundance (Elvidge, Glenny et al. 2006; Warnecke, Weidemann et al. 2008). In these studies the

major findings were that: 1) the number of genes regulated by the HIF pathway is limited to a few hundred; 2) in the examined cells most of the hypoxic responses are mediated via HIF-1α; 3) the capacity of HIF-2α to transactivate expression is limited to a small number of genes in the cell lines tested and 4) the concordance between hypoxia and DMOG in terms of regulation of transcription is very high (MCF-7 cells in (Elvidge, Glenny et al. 2006) and Hep3B cells in (Warnecke, Weidemann et al. 2008)). A number of the genes which were strongly regulated by hypoxia had already been described (by conventional techniques) as HIF targets (e.g. *EPO*, *GLUT1*, and *CA9*) with functional HIF-binding sites predominantly within promoter regions. However, these conventional promoter-biased techniques will underestimate the true number of direct transcriptional targets of HIF, since binding may elicit gene regulation over considerable genomic distances. For example, both the *EPO* and *EGLN3* HIF-binding sites are distant from the promoter. Conversely, whilst Elvidge et al. have reported similar numbers of genes that are up- or downregulated by hypoxia, reports of HIF-binding at downregulated gene are sparse in comparison to upregulated loci suggesting either a different pattern of binding or indirect regulation (Elvidge, Glenny et al. 2006).

Since the degree to which HIF regulated genes are primary or secondary transcriptional or post-transcriptional targets of the HIF pathway remains unclear I resolved to define exactly where HIF is binding in the human genome, whether there are differences between the two HIF-α isoforms and how this correlates to HIF dependent gene regulation. For this I used chromatin immunoprecipitation using antibodies directed against HIF-1α, HIF-2α and HIF-1β in MCF-7 cells coupled to high-throughput sequencing (ChIP-seq).

2.2 Results

2.2.1 ChIP and ChIP-seq for HIF-α subunits

In this part of the study MCF-7 breast cancer cells were examined for HIF-binding. HIF-α subunits were stabilised using either 2 mM dimethyloxalylglycine (DMOG) or hypoxia (0.5%) for 16 h. Figure 2.1a shows control western blots confirming stabilisation of HIF-α sub-units by DMOG and induction of target genes known to be induced in this cell line (Glut1, Ca9). For ChIP experiments proteins were cross linked with DNA using 1% formaldehyde and, after cell lysis, chromatin was sheared by sonication using a water bath sonicator to achieve a fragment size of approximately 500 bp (Figure 2.1b). Anti-sera against HIF-1α and HIF-2α were employed for immunoprecipitation. These antibodies had previously been generated and validated in the laboratory. Their use for ChIP experiments examining HIF-binding at the *CA9* and the *EGLN3* locus was described previously ((Lau, Tian et al. 2007), Figure 2.1c) and they have been used successfully for ChIP-on-chip studies (Mole, Blancher et al. 2009). Chromatin immunoprecipitations using rabbit preimmune serum were used as negative controls. For ChIP-seq experiments an aliquot (15 µl of 25 µl total) of captured DNA fragments from one of the replicates for each anti-serum described in Mole et al. was forwarded to the Genomic sequencing facility at the Wellcome Trust Centre and sequencing was performed as described in the methods section (Mole, Blancher et al. 2009). 36 bp reads were mapped to the human genome (version HG 18). Two sample analysis (HIF-α antibody vs control serum), subsequent assessment, peak finding, motif analysis and mapping of RCGTG motifs were undertaken using the CisGenome software suite (http://www.biostat.jhsph.edu/~hji/cisgenome). Data were also uploaded to the UCSC

Genome Browser for viewing (http://genome.ucsc.edu/). Examples of computed sequencing data from the CisGenome browser (Figure 2.1d) and the UCSC Genome Browser (Figure 2.2) illustrating read densities at the *EGLN3* HIF-binding locus are shown.

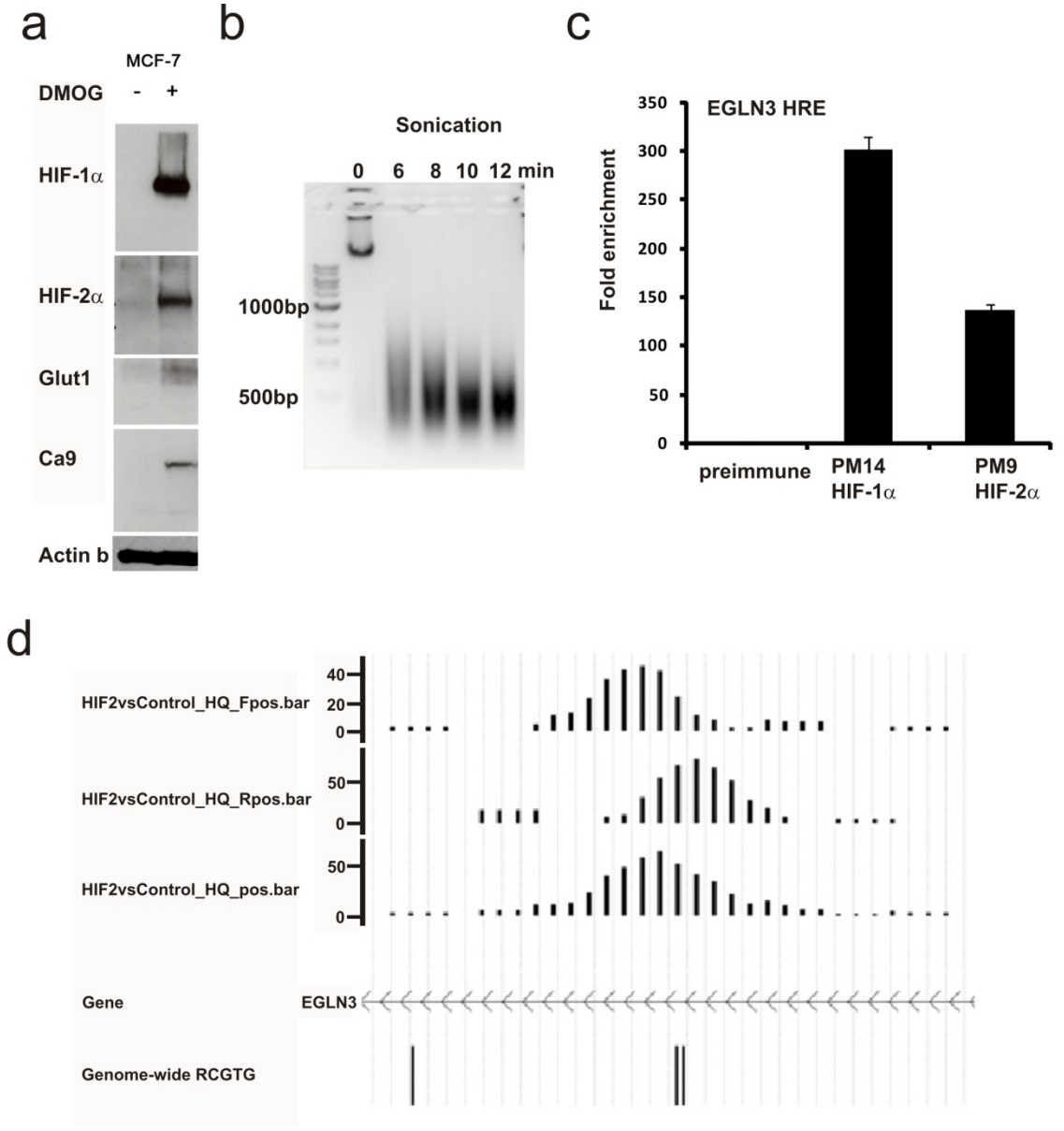


Figure 2.1: (a) Western blot using MCF-7 lysates to confirm stabilisation of HIF-α subunits by 2 mM DMOG and induction of HIF target genes. (b) Agarose gel showing the effect of sonication for different length of time on the size of DNA fragments. DNA was isolated from the sheared chromatin after reversion of the protein-DNA cross linking by heat. Routinely between 12 to 15 min of sonication were used and fragment size was checked on a gel. (c) ChIP experiment in MCF-7 cells exposed to 2 mM DMOG for 16 h using preimmune, anti-HIF-1α or anti-HIF-2α serum. qPCR results in triplicates from one representative ChIP experiment using primers spanning the *EGLN3* intronic HRE were normalised to values for beta actin and preimmune serum using the ΔΔct method. Bars show mean ± s.d. (d) ChIP-seq tracks from CisGenome showing number of reads per 25 base pair window in forward (top track) and reverse (middle track) direction and the sum of both (bottom track) at the *EGLN3* intronic HRE (bottom lane) for HIF-2α after normalisation to the preimmune control.

Inspection of the read density achieved with HIF-ChIP at the previously described HIF-binding site in the first intron of *EGLN3* (Pescador, Cuevas et al. 2005) shows well-defined peaks encompassing the HRE for HIF-1α and HIF-2α (Figure 2.2). This confirms the high resolution of binding sites achieved by ChIP-seq which allows defining the binding event within approximately 200 bp resolution. Also the concordant results from both data sets confirm previous results that HIF-1α and HIF-2α can bind to the same HREs (Lau, Tian et al. 2007; Mole, Blancher et al. 2009).

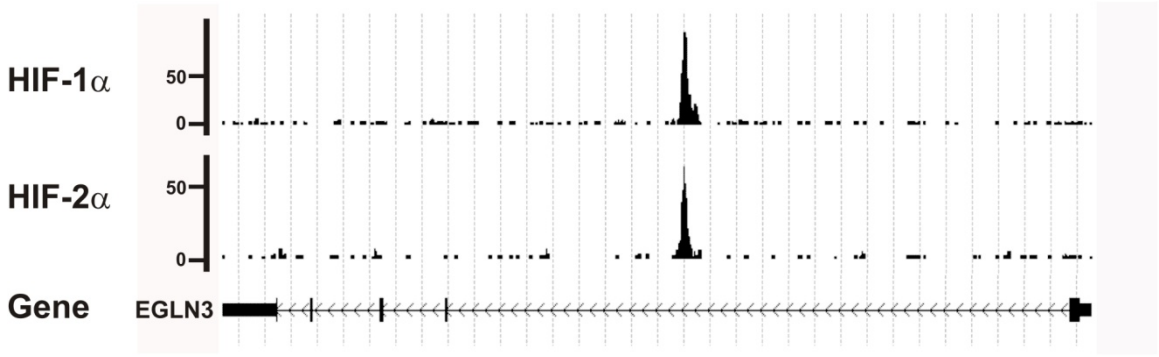


Figure 2.2: Read density tracks (UCSC Genome Browser) from ChIP-seq experiments for HIF-1α and HIF-2α in MCF-7 cells across the entire *EGLN3* gene. HIF-binding is detected at the previously described intronic HRE site.

2.2.2 Genome-wide HIF-1α and HIF-2α binding

An important challenge with large ChIP-seq datasets is to identify an appropriate threshold to define significant peaks. Using the gene loci identified by previous ChIP-on-chip analysis of MCF-7 human breast cancer cells treated with dimethyloxalylglycine (DMOG) for 16 hours (Mole, Blancher et al. 2009) as a "seed" data set, the effect of altering the threshold used to define ChIP-seq peaks, on the degree of overlap between the two data sets was examined. This identified a threshold of approximately 20 reads per window for HIF-1α (and 25 reads per window for HIF-2α) in the ChIP-seq experiment below which the overlap with the ChIP-on-chip data fell to background (Figure 2.3).

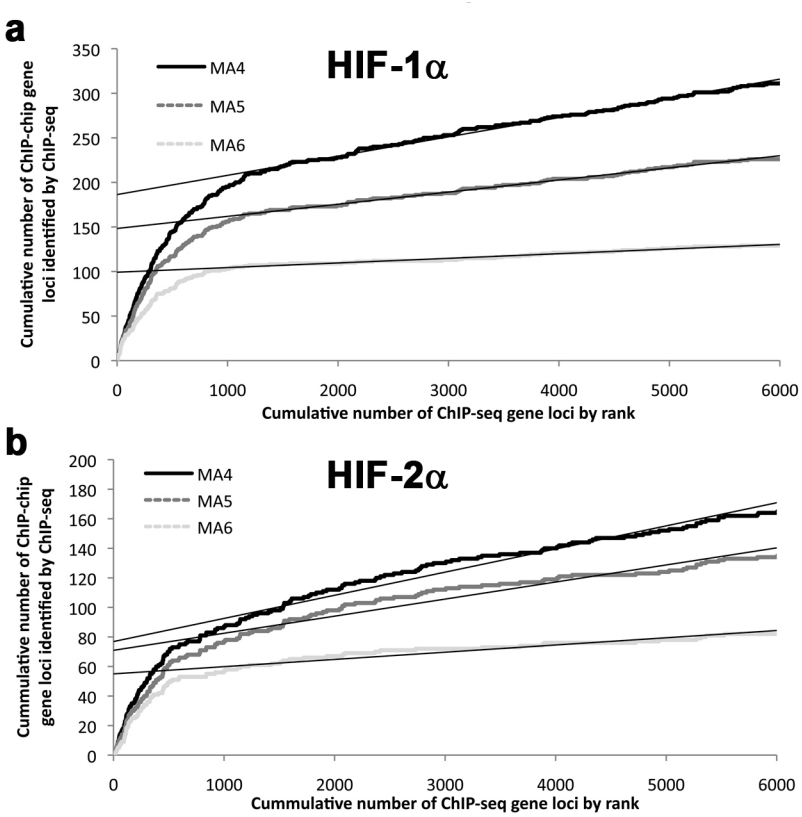


Figure 2.3: Gene loci identified in the HIF-1α and HIF-2α ChIP-seq analyses were ranked according to the maximum height of the ChIP-seq signal. Using ChIP-chip gene sets (of varying stringency - defined by the denoted z-score thresholds (moving average: MA)), the cumulative number of gene loci identified in both the ChIP-seq and ChIP-chip datasets was identified, as the ChIP-seq threshold was reduced, and was plotted against the rank in the ChIP-seq analysis. Significant overlap was seen between the ChIP-seq and ChIP-chip datasets amongst the highest 1000 ChIP-seq peaks for both HIF-1α (a) and HIF-2α (b). Beyond this number the overlap fell to statistical background. The threshold (point of inflection) at which this happened was robust to the stringency with which the ChIP-chip peaks were defined. Using this threshold together with additional filters, high stringency gene sets were identified for both HIF-1α and HIF-2α ChIP-seq analyses.

To define high stringency ChIP-seq binding sites more rigorously these peak height thresholds were used in addition to three further filters. Firstly, the number of reads in each direction must comprise at least one third of the total (i.e. the peak height for both the forward and the reverse strands must be greater than 7). Secondly, the estimated false discovery rate as calculated by the CisGenome software must be less than 0.05. Finally gene loci with more than 4 binding sites were excluded as manual analysis revealed these to be due to regions of high background, mapping to repeat elements or contamination with expression plasmids.

This analysis identified 400 high stringency HIF-1α binding sites at 356 gene loci (Appendix 8.2.1) and 425 high stringency HIF-2α binding sites at 357 gene loci (Appendix 8.2.2). Ranking the binding gene loci of one isoform by height and testing the overlap with the high stringency loci of the other isoform the overlap of binding sites between the two HIF-α subunits was determined to be approximately 250 binding loci (Figure 2.4). The use of quantitative PCR analysis to independently assess the veracity of a selection of the peaks identified by HIF-1α ChIP-seq, confirmed a very high rate of true positivity and the overlap between HIF-1α and HIF-2α (Figure 2.5).

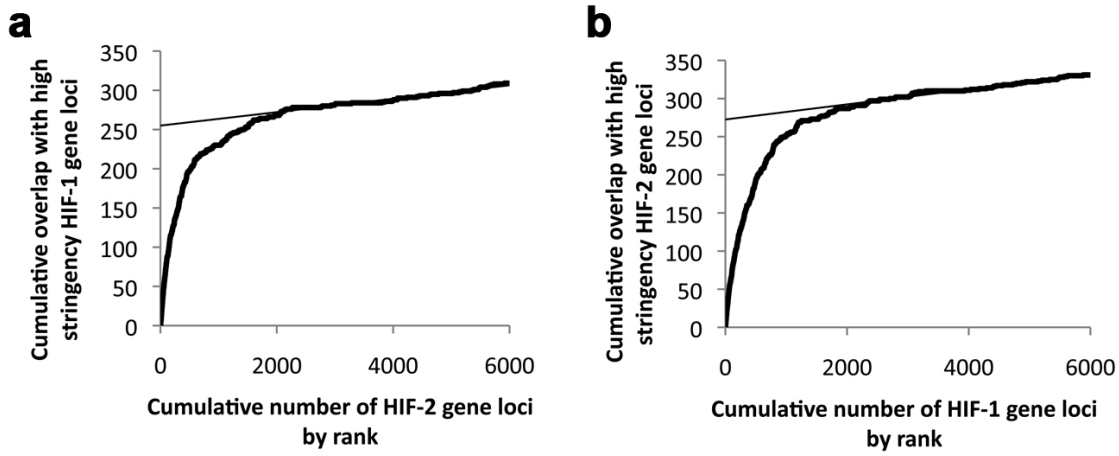


Figure 2.4: (a) The overlap between the set of HIF-1α binding gene loci defined by these high stringency criteria and all HIF-2α peaks was examined as the threshold for definition of HIF-2α ChIP-seq peaks was reduced (similar to analysis in Figure 2.3) and again showed a high degree of overlap amongst the 1000 highest HIF-2α peaks corresponding to approximately 250 common gene loci. (b) The reverse analysis using the high-stringency HIF-2α binding gene loci and all HIF-1α binding loci confirmed a comparable degree of overlap.

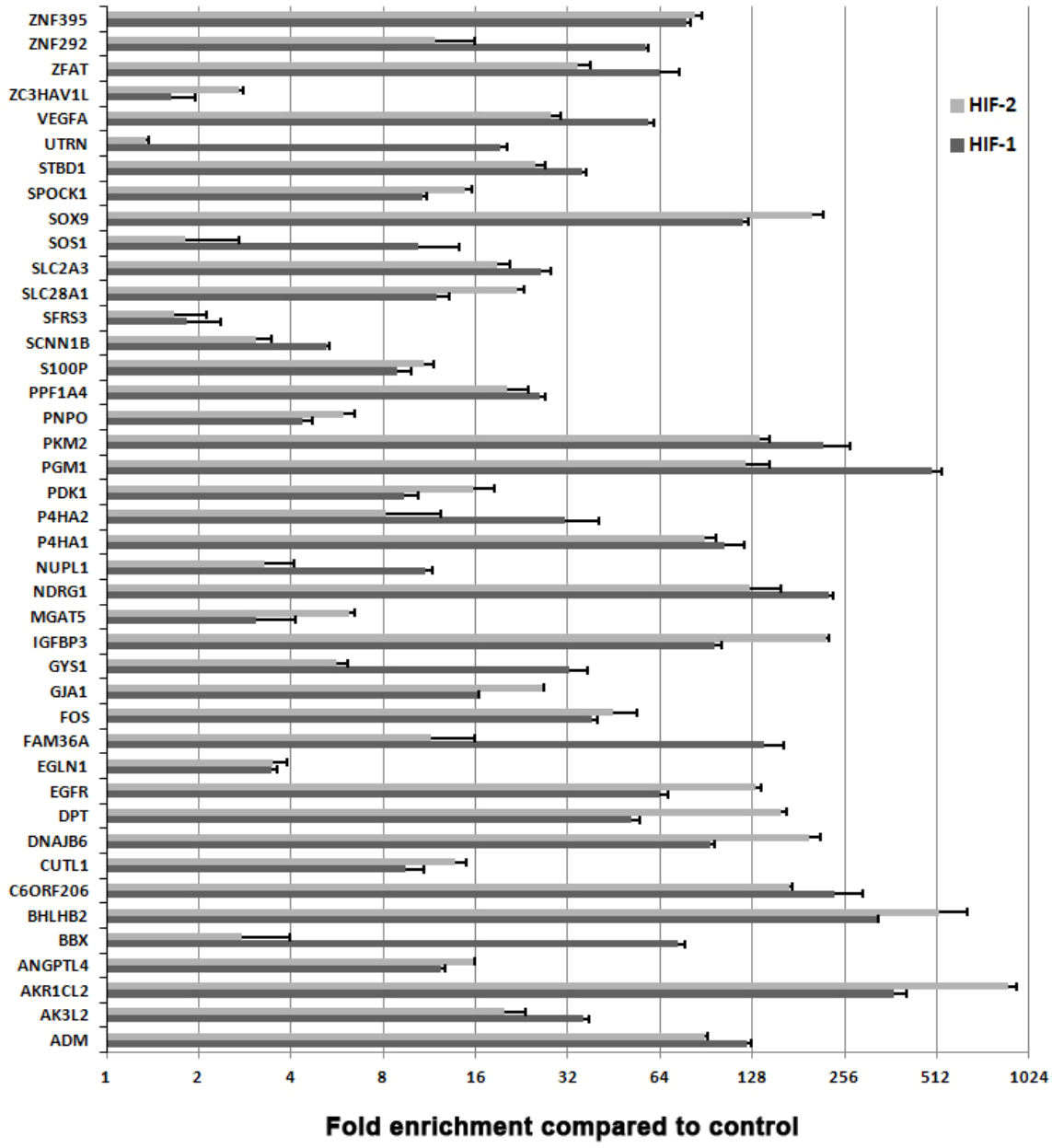


Figure 2.5: Confirmation of HIF-1α and HIF-2α binding by ChIP qPCR at a selection of peaks annotated to the next nearest gene. Bars show mean ± s.d. (n=3).

2.2.3 HIF-1β binding in MCF-7 cells

As an independent check on HIF-α binding sites a separate ChIP-seq analysis of HIF-1β DNA-binding sites using cells treated with 0.5% hypoxia or normoxia for 16 hours was undertaken. This allowed the concordance of HIF-binding under hypoxic or DMOG treated conditions, respectively, to be examined. This has previously been described to be very high in terms of transcriptional regulation (Elvidge, Glenny et al. 2006; Warnecke,

Weidemann et al. 2008). It also allows screening for potential normoxic HIF-1β binding events independent of the presence of HIF-α subunits. A commercial HIF-1β antibody was used for ChIP experiments in lysates from MCF-7 cells exposed to normoxia or hypoxia. In these experiments strong enrichment for HIF-1β binding induced by hypoxia was seen at the *EGLN3* HRE, the *SAP30* HRE and the *MLL5* HRE confirming the suitability of this anti-serum for ChIP experiments (Figure 2.6) and suitability of the material for next-generation sequencing.

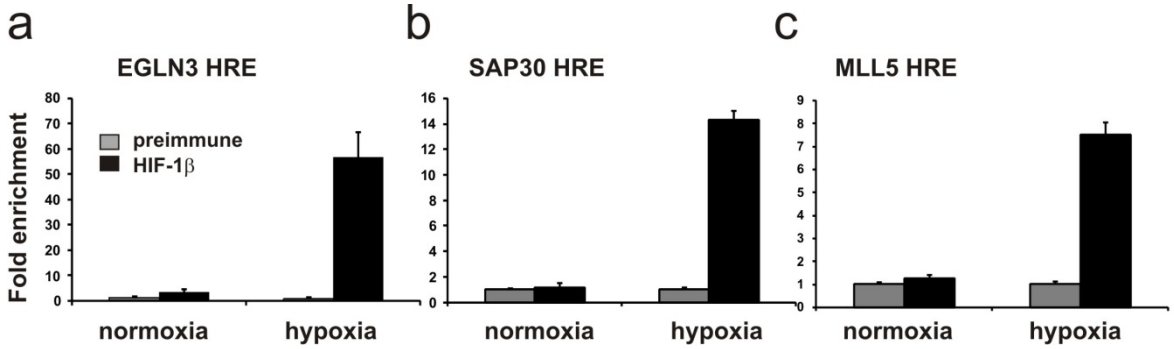


Figure 2.6: ChIP qPCR for (a) the *EGLN3* HRE, (b) the *SAP30* HRE and (c) the *MLL5* HRE in MCF-7 cells exposed normoxia or hypoxia 0.5% for 16 h using preimmune serum or an anti-serum against HIF-1β. Bars show mean ± s.d. (n=3). Note that binding of HIF-1β is highly inducible by hypoxia at all three sites tested.

DNA samples for the four immunoprecipitations (preimmune serum and HIF-1β anti-serum for normoxic or hypoxic conditions) were submitted to the sequencing facility and whole genome sequencing was performed as described. After mapping of the reads and uploading of the data set for visualisation, inspection of the *EGLN3* HRE locus again revealed a well defined peak for the HIF-1β subunit in hypoxia overlapping signals from the HIF-α isoforms (Figure 2.7).

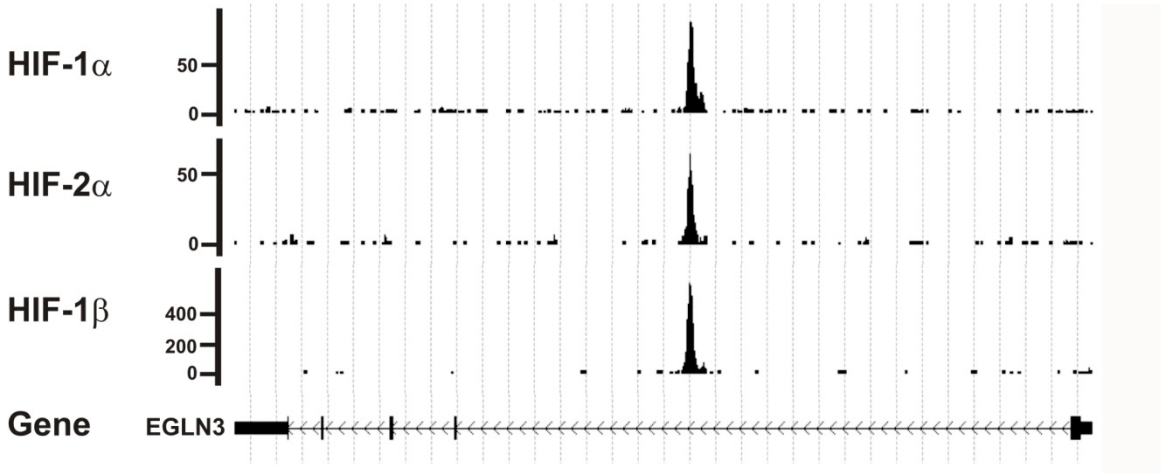


Figure 2.7: Read density tracks from ChIP-seq experiments for HIF-1α, HIF-2α and HIF-1β in MCF-7 cells across the entire *EGLN3* gene. HIF-binding is detected at the previously described intronic HRE site. HIF-α subunits were stabilised using 2 mM DMOG, HIF-1β binding was induced by hypoxia (0.5% O_2) for 16 h. Note that the accumulated number of reads is higher in the HIF-1β experiment due to a higher number of total reads in this experiment.

To further check validity of the results, other known HIF-binding sites were examined. Striking overlap of hypoxic HIF-1β binding was seen with HIF-α binding events induced by DMOG at several gene loci (Figure 2.8). Furthermore, in the control ChIP experiments (preimmune serum in normoxia and hypoxia, HIF-1β in normoxia) no significant signal could be detected in these regions. This implies that HIF-1β binding like HIF-α binding is highly inducible by hypoxia and though HIF-1β is present it does not bind at these loci under normoxic conditions.

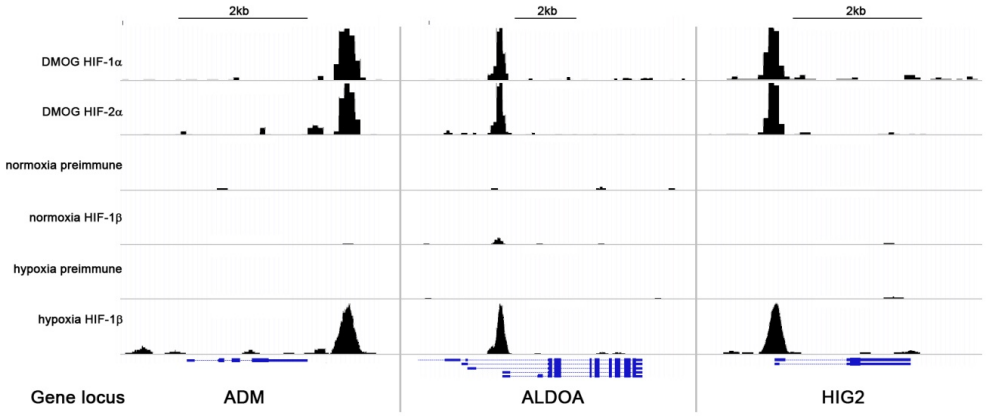


Figure 2.8: Read densities for HIF-binding at promoters of the known HIF-target genes *ADM*, *ALDOA* and *HIG2*. Hypoxic HIF-1β binding is concordant with DMOG-induced HIF-α binding. Note the low background signals in control ChIP-seq experiments (normoxia preimmune, normoxia HIF-1β and hypoxia preimmune).

To examine HIF-1β binding on a genome-wide level hypoxic HIF-1β signals were ranked, by CisGenome, according to peak height, and the degree of overlap with the high stringency HIF-1α and HIF-2α binding sites was examined. This shows a high degree of overlap between HIF-α and HIF-β binding sites, falling to background levels after the top 2000 HIF-1β binding sites (Figure 2.9a, c). HIF-1β binding was seen at 356 out of 400 HIF-1α sites and 301 out of 425 HIF-2α binding sites (Figure 2.9a, c).

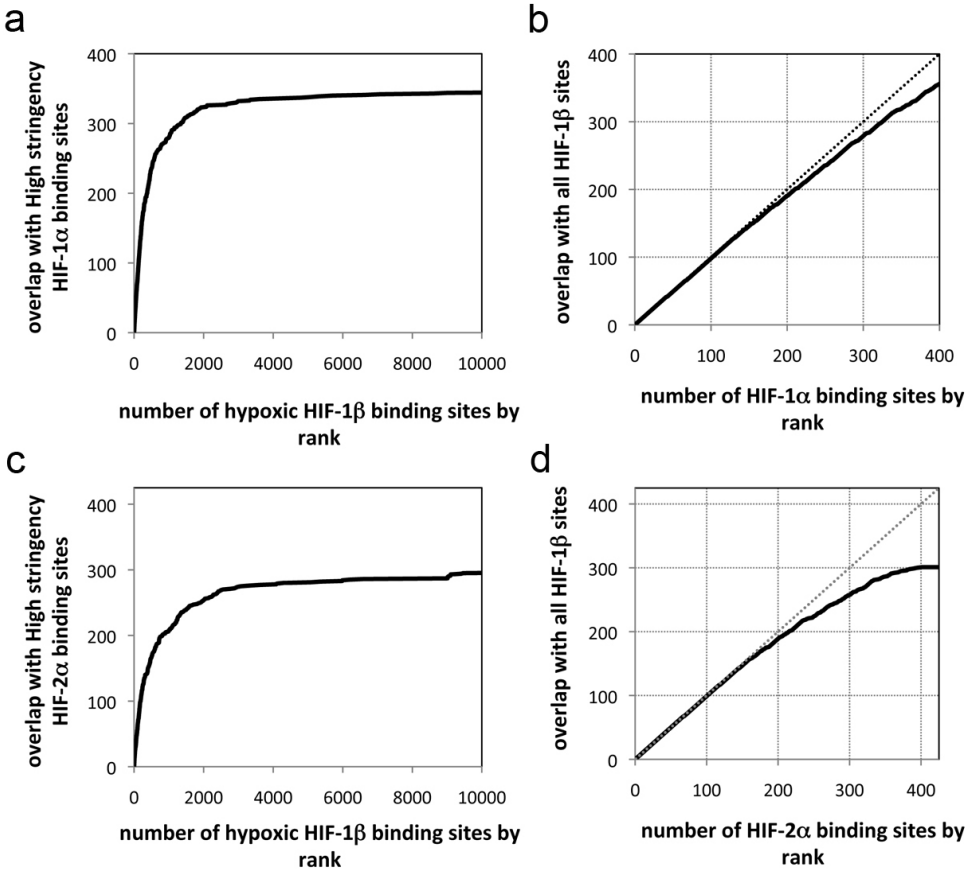


Figure 2.9: (a) Hypoxic HIF-1β ChIP-seq peaks were ranked according to peak height and the number overlapping with the high-stringency HIF-1α binding sites was plotted. (b) The high-stringency HIF-1α binding sites were ranked according to peak height and the number overlapping with the top 10,000 hypoxic HIF-1β ChIP-seq peaks was plotted. (c and d) The analysis was repeated for HIF-2α and hypoxic HIF-1β.

However, most of the non-overlap was accounted for by the weaker HIF-α peaks within each group. Indeed, when only the top 200 HIF-1α and HIF-2α sites were considered the overlap rates with HIF-1β were 95% and 94.5% respectively (Figure 2.9 b, d). This implies

that the non-overlap is resulting from an increasing degree of false positivity amongst the HIF-α datasets as the threshold is approached. It also indicates either that HIF-α can only bind DNA as a heterodimer with HIF-1β or that if HIF-α is able to bind DNA with an alternative partner that it does so at the same sites as the HIF-α/HIF-1β heterodimer or with much lower affinity/occupancy. The reverse analysis, using the top 200 and 500 HIF-1β sites shows overlap rates with combined HIF-α sites of 95% and 87% respectively, again indicating that essentially all hypoxic HIF-1β sites bind HIF-α (Figure 2.10). Furthermore, the high degree of concordance observed between DMOG induced HIF-α binding sites and hypoxia induced HIF-1β sites suggests that the two treatments are comparable with respect to the pattern of HIF DNA-binding that they induce.

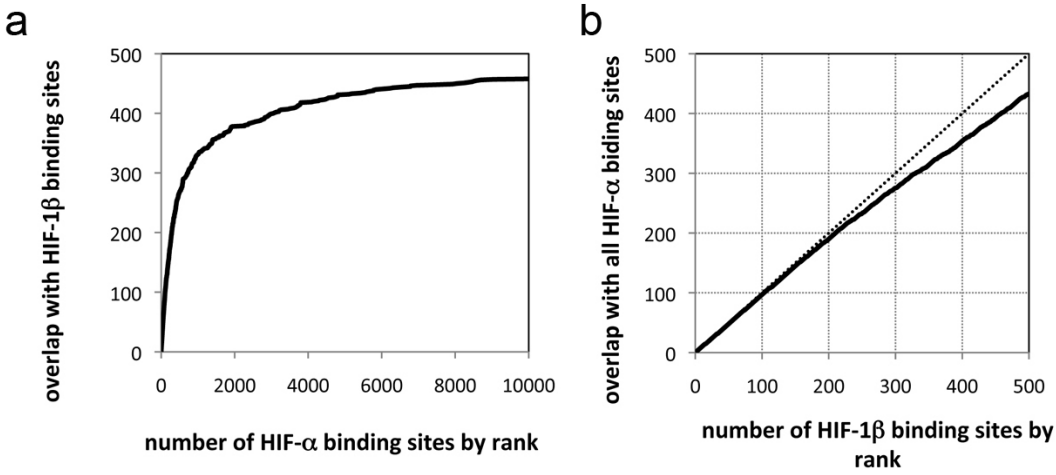


Figure 2.10: (a) HIF-α ChIP-seq peaks were ranked according to peak height and the number overlapping with the top 500 hypoxic HIF-1β binding sites was plotted. (b) The top 500 hypoxic HIF-1β binding sites were ranked according to peak height and the number overlapping with the top 10,000 HIF-α ChIP-seq peaks was plotted.

2.2.4 Distribution of HIF-binding in the vicinity of transcriptional start sites

The ability to identify HIF-binding sites, independently of the coverage afforded by tiled promoter microarrays, permits true genome-wide analysis of HIF-α DNA-binding, thus enabling a complete description of the distribution of HIF-binding sites. Of the 400 high

stringency HIF-1α binding sites identified by ChIP-seq above, 201 (50%) were novel sites, distant from promoter regions (as defined by the Human Promoter 1.0R microarray) whereas 316 (74%) of 425 HIF-2α binding sites lay outside these regions. These differing proportions likely account for the relatively greater number of novel HIF-2α binding sites identified by ChIP-seq compared to Chip-on-chip and suggest differing distributions of the two isoforms with respect to the promoter.

To test this, the frequency distribution of HIF-1α and HIF-2α binding sites around the nearest transcriptional start site (TSS) was examined. Using 5 kb windows, we observed clustering of both HIF-1α and HIF-2α binding sites close to the TSS, with a low frequency of binding extending to much greater distances. However, HIF-1α was more likely to bind close to the TSS, with approximately 40% of HIF-1α binding sites lying within 2.5 kb of the TSS, as opposed to approximately 20% of HIF-2α binding sites (Figure 2.11).

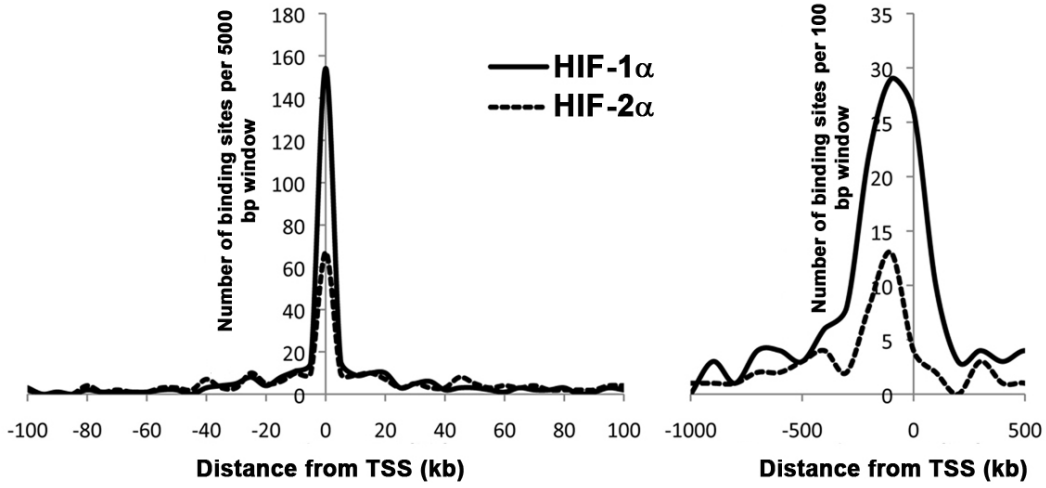


Figure 2.11: Distribution of HIF-binding sites around the transcriptional start site (TSS). The position of HIF-1 and HIF-2 binding sites with respect to the nearest TSS was plotted as a frequency distribution using (left) 5 kb windows and (right) 100 bp windows.

More detailed analysis of the binding patterns using 100 bp windows indicated that for either HIF-α isoform, the most frequent position of the binding site laid approximately 100 bp 5' to the TSS, within the core promoter (Figure 2.11 right). Figure 2.12 shows the distribution of HIF-binding sites with respect to gene architecture stratified in intergenic, 5'-UTR, 3'-UTR, intronic and exonic regions.

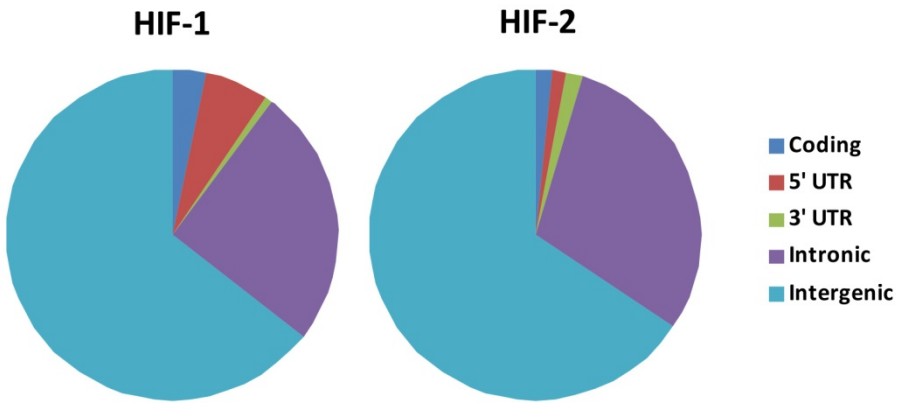


Figure 2.12: Distribution of HIF-binding in respect to gene architecture.

2.2.5 Distribution of HIF-binding across the genome

The full genome-wide coverage afforded by ChIP-seq also permits analysis of clustering of HIFα binding sites across the genome. Analysing nearest-neighbour distance, for both HIF-1α and HIF-2α binding loci, revealed a non-Poisson distribution. However, the spread was comparable with the distribution of identical numbers of randomly selected genes, indicating that HIF-1α and HIF-2α target genes are randomly distributed amongst known transcripts (Figure 2.13). The same distribution was observed when examining the distribution of HIF-regulated genes as defined by Elvidge et al. This indicates that there are no clusters or "hot spots" of genes in the genome preferentially activated by HIF.

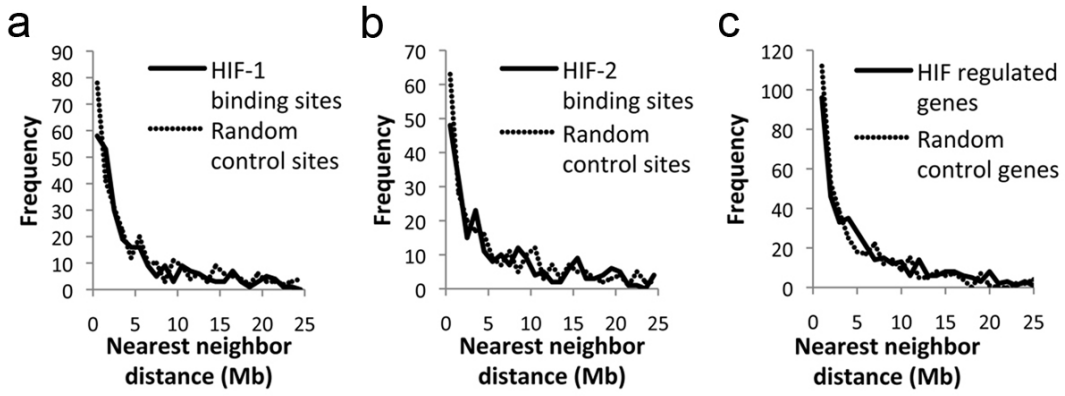


Figure 2.13: Frequency distribution of the nearest-neighbour distance between (a) HIF-1-binding gene loci, (b) HIF-2-binding gene loci, and (c) HIF-regulated gene loci. In each case, the frequency distribution of nearest-neighbour distances for an equal number of randomly selected control gene loci is shown for comparison.

2.2.6 Pathway analysis

There is a high overlap between HIF-1 and HIF-2 binding loci. However, recent reports point to differences in transcriptional activity of the isoforms (Hu, Sataur et al. 2007; Lau, Tian et al. 2007). To test whether these differences occur on the level of DNA-binding the HIF-binding loci (as defined by the next nearest gene) were examined using the Ingenuity Pathways Analysis software tool (Version IPA 8.8-3204). Twenty canonical pathways were statistically over-represented among HIF-binding loci with several demonstrating bias toward either HIF-1 or HIF-2 binding (Figure 2.14). Interestingly, the OCT-4 pathway has been described to be affected more by HIF-2α than HIF-1α indicating that although a high degree of overlap exists between HIF-1 and HIF-2 binding some binding loci are preferring one over the other isoform.

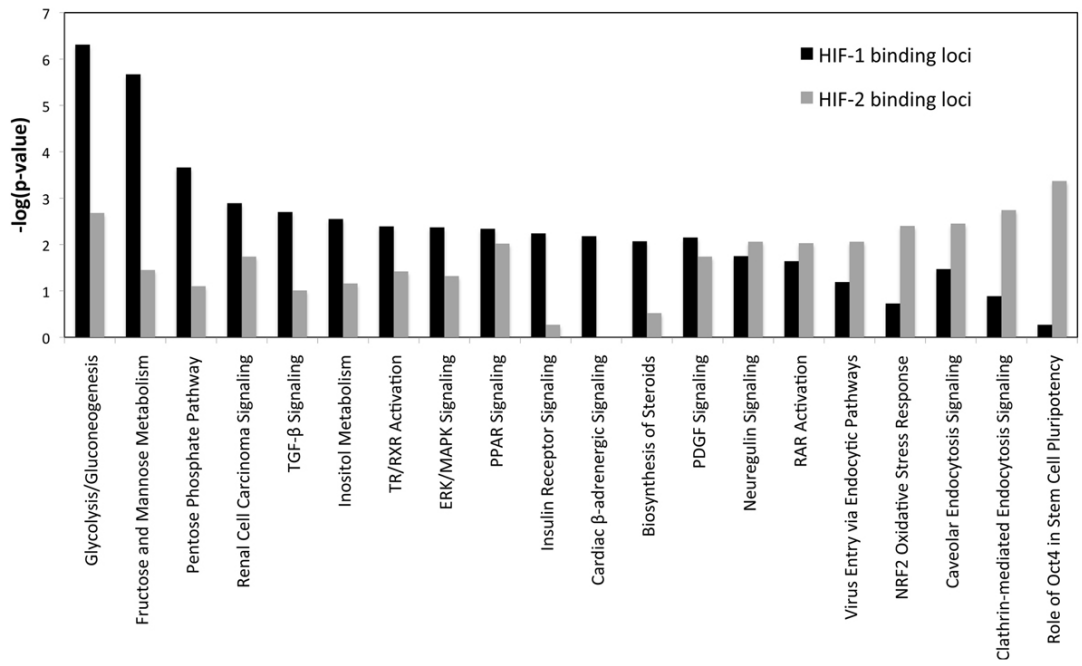


Figure 2.14: The Ingenuity Systems Inc (IPA Version 8.8-3204) software tool was used to examine the HIF-1 and HIF-2 binding loci for functional pathways. Canonical pathways statistically over-represented among HIF-binding loci are shown together with the respective -log10 (*P* value) for each isoform.

2.2.7 Analysis of transcription factor binding site motifs at HIF DNA-binding sites

Based on the analysis of around 50 HIF-regulated genes, the minimal cis-regulatory element required for HIF-dependent transactivation has been defined as the RCGTG motif (Wenger, Stiehl et al. 2005). The unbiased pan-genomic identification of in excess of 400 HIF-binding sites, together with their more precise localization, has allowed a more definitive statement regarding sequence preferences surrounding this core motif. To do this, two independent approaches were employed. Firstly, in an unbiased analysis, high stringency HIF-α binding sites were examined for common novel motifs using the Gibbs Motif Sampler module of CisGenome. Novel motifs identified in this way were then examined for enrichment by determining the relative number of occurrences in the ChIP-seq defined HIF-binding regions compared to a control set of regions. As expected, the most enriched motif identified in this way for both HIF-1α and HIF-2α binding regions contained the core RCGTG motif with a preference for A over G at the R position (Figure

2.15). However, these motifs also exhibited extended sequence preferences both 5' and 3' to this core motif. In both cases preference was seen for T immediately 5' to the RCGTG and for C immediately 3' to this motif. The motif identified amongst HIF-1α binding regions also identified a preference for C or G 2-bp 5' to the R position. In addition, this approach also identified a number of other motifs that occur more often in regions of HIF-binding than in control regions, including an AP-1 like motif.

a

Motif	r1	r2	r3	Occurrences
	7.5	16.2	18.3	291
	3.1	4.8	5.4	132
	3.1	3.1	3.6	255

b

Motif	r1	r2	r3	Occurrences
	7.9	14.3	19.5	255
	4.2	4.2	5.8	154
	3.9	6.7	9.3	51

Figure 2.15: HIF-1 (a) and HIF-2 (b) binding regions were examined for commonly occurring novel motifs using the Gibbs Motif Sampler module of CisGenome. Enrichment of motifs identified in this way was then examined by determining the relative occurrence of each motif in HIF-binding regions compared to a control set of regions (r1), of phylogenetically conserved motifs within phylogenetically conserved portions of HIF-binding regions compared to control regions (r2) and of phylogenetically conserved motifs within all portions of the HIF-1 binding and control regions (r3). The sequence logos for motifs with all 3 enrichment ratios greater than 2 are shown. The height of each letter is proportionate to its frequency, with the most frequent at the top. The height of each stack is then adjusted to signify the information content of the sequences at that position (measured in bits).

Whilst the above approach will determine novel motifs amongst sets of binding regions, it is poor at distinguishing the relative abundance of bases in two related motifs, for instance sequences binding HIF-1α versus HIF-2α. Therefore, in a second approach, predicated on the RCGTG motif, sequences 100 bp up and downstream of each unique

HIF-binding RCGTG motif identified in the HIF-1α and HIF-2α ChIP-seq datasets were obtained. In order to confidently identify the RCGTG motif responsible for HIF-binding, only gene loci at which a single RCGTG motif (or two overlapping RCGTG motifs) was present in the defined HIF-binding region were analysed (Figure 2.16). Positions at which the base frequency differs significantly from the background distribution are shown highlighted in grey ($p<0.05$, with Bonferroni correction). Both HIF-1α and HIF-2α binding regions are relatively GC rich, reflecting their concentration within promoter regions. For both HIF-1α and HIF-2α an extended HRE consensus motif is identified, which is remarkably similar for the two isoforms. Indeed, no significant difference in base frequency was observed between the HIF-1α and HIF-2α binding regions at any position flanking the RCGTG motifs. Furthermore, this motif was remarkably similar for binding sites both close to and distant from the transcriptional start site. At position 0 (the position of R), preference is seen for A over G. Preference is also seen for T at position -1. Importantly, C is strikingly under-represented at position -1, implying that the CACGTG E-box motif is inhibited from binding HIF. This exclusion may result from competition by other proteins that recognize the E-box and reduce their availability to bind HIF, including potentially the HIF-1β homodimer. Outside the core RCGTG motif preference is seen for a CAC motif between positions +13 and +15. Mutation of this latter region has previously been shown to affect functional activity of the erythropoietin 3' enhancer. Interestingly, this CAC motif occurs at an interval from the core RCGTG motif of approximately one revolution of the DNA double helix, potentially bringing these motifs into juxtaposition.

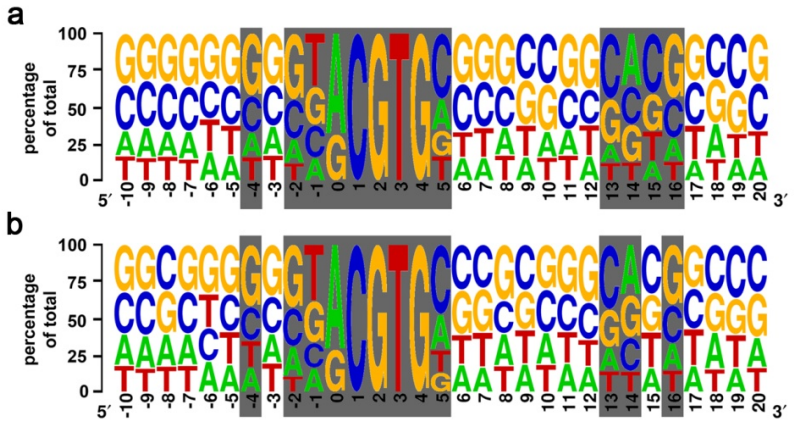


Figure 2.16: HIF-1 (a) and HIF-2 (b) binding sites containing a single RCGTG motif were analysed for sequence conservation beyond this core motif. Sequences were extended for 100 bp upstream and downstream of the RCGTG motif and aligned. At each position, the observed base frequency was compared with the expected using a Chi-squared test. The relative occurrence of each nucleotide at each position is indicated as a percentage of the total by the letter height. Positions showing non-random base frequency (p<0.01 after Bonferroni correction) are highlighted in grey.

2.2.8 Correlation of HIF-binding with DNAse1 hypersensitivity

Within the reference genome CisGenome identified 1,114,925 RCGTGs of which only 632 (0.06%) were identified in HIF-1α and 502 (0.05%) in HIF-2α binding regions. Even allowing for false negative results arising from the high stringency criteria used in the definition of these binding sites, the large majority of RCGTG motifs do not bind HIF. With the exception of excluding some E-box motifs, sequence analysis does not provide any definite rules to define potential binding sites further. To further determine the extent to which epigenetic mechanisms, in particular chromatin accessibility, as defined by DNAse1 hypersensitivity, might contribute to the definition of a HIF-binding site a correlation of the HIF-binding results with normoxic DNAse1 hypersensitivity peaks from the ENCODE consortium dataset 1 for Hypersensitivity by Digital DNAse1 in MCF-7 cells (wgEncodeUwDnaseSeqPeaksRep1Mcf7) was performed.

Overall, 288 (72%) of the HIF-1α binding sites and 287 (68%) HIF-2α binding sites overlapped with DNAse1 hypersensitive peaks and HIF-binding regions demonstrated a

much higher concentration of hypersensitive peaks than the immediate flanking regions. When all of the 1,114,925 RCGTG motifs, genome-wide, are considered DNAse1 hypersensitive RCGTG motifs are 32 times as likely to bind HIF-1α and 41 times as likely to bind HIF-2α as insensitive motifs (Figure 2.17). This suggests that DNAse1 hypersensitivity in normoxic cells would be an important predictor of HIF-α binding to its consensus recognition site.

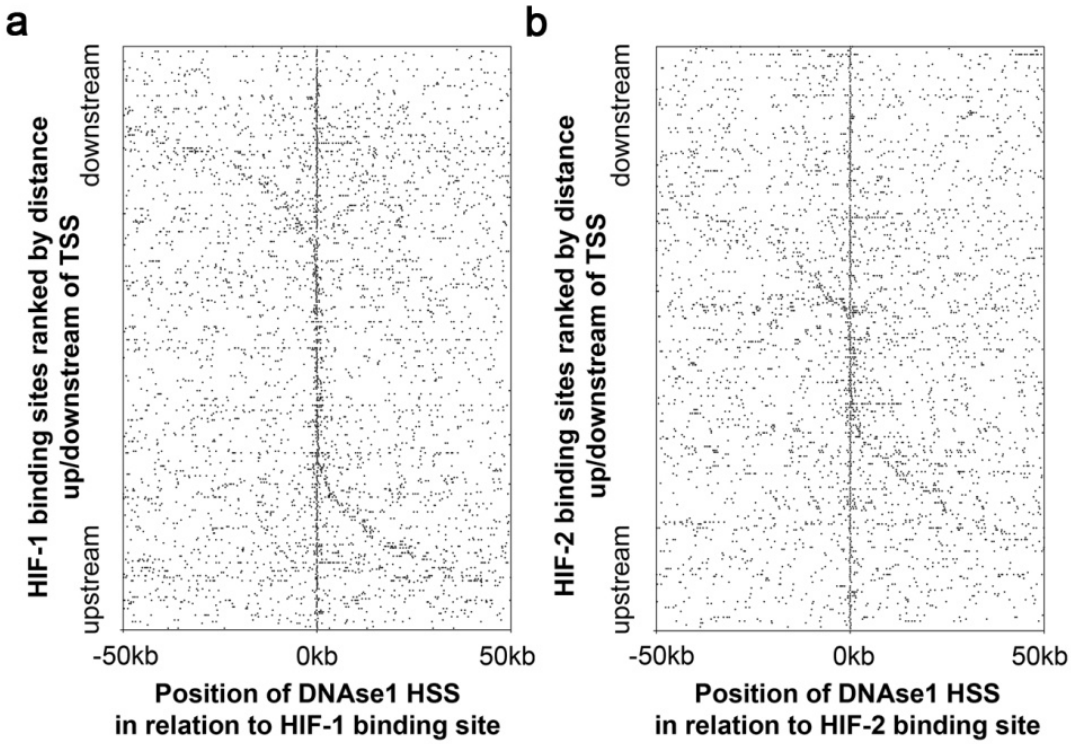


Figure 2.17: HIF-1 (a) and HIF-2 (b) binding sites were ranked according to distance up/downstream of the nearest TSS (vertical axis). Normoxic DNAse1 hypersensitivity peaks were identified for MCF-7 cells using the ENCODE consortium dataset 1 for Hypersensitivity by Digital DNAse1. For each HIF-1 binding site, the position of each DNAse1 hypersensitive peak (within 50 kb), relative to the centre of the HIF-binding region, was plotted (horizontal axis). This analysis shows two distinct patterns of enrichment. The first (Z-shaped) clustering of hypersensitive peaks corresponds to enrichment of hypersensitive sites within the core promoters of the neighbouring genes. The second pattern (centred on 0 kb) corresponds to enrichment of hypersensitive peaks at the HIF-binding sites themselves. This second pattern of enrichment of hypersensitive peaks was seen across all HIF-binding sites, irrespective of their distance from the TSS, further confirming the validity of these more distant HIF-binding sites.

To test this at high resolution and to explore the extent to which co-incidence between DNAse1 hypersensitivity and HIF-binding was observed away from promoter regions (where DNAse1 hypersensitivity sites may be predicted to be enriched in any case), the

HIF-binding loci defined by ChIP-seq were ranked according to distance up/downstream of the nearest TSS. Then this rank was plotted against the distance from the centre of the HIF-binding region (defined experimentally by ChIP-seq) of either all RCGTG motifs, or just the RCGTG motifs that coincided with DNAse1 hypersensitivity sites, (Figure 2.18). An expectation of these analyses is that correctly identified HIF-binding RCGTG motifs will align at the centre of the region defined by ChIP-seq whereas RCGTG motifs that are not binding HIF will be randomly distributed. Figure 2.18a shows the plot for all RCGTGs and reveals the expected alignment against a background of 'noise' from the non-HIF binding RCGTGs. When the data is restricted to RCGTG motifs that fall within DNAse1 hypersensitive sites then a much-improved 'signal-to-noise' ratio is observed (Figure 2.18b) consistent with the predictive power of DNAse1 hypersensitivity. Importantly, this is observed whatever the rank distance of the HIF-binding sites from the nearest TSS, indicating that normoxic DNAse1 hypersensitivity is a strong determinant of HIF-binding even at sites remote from promoters.

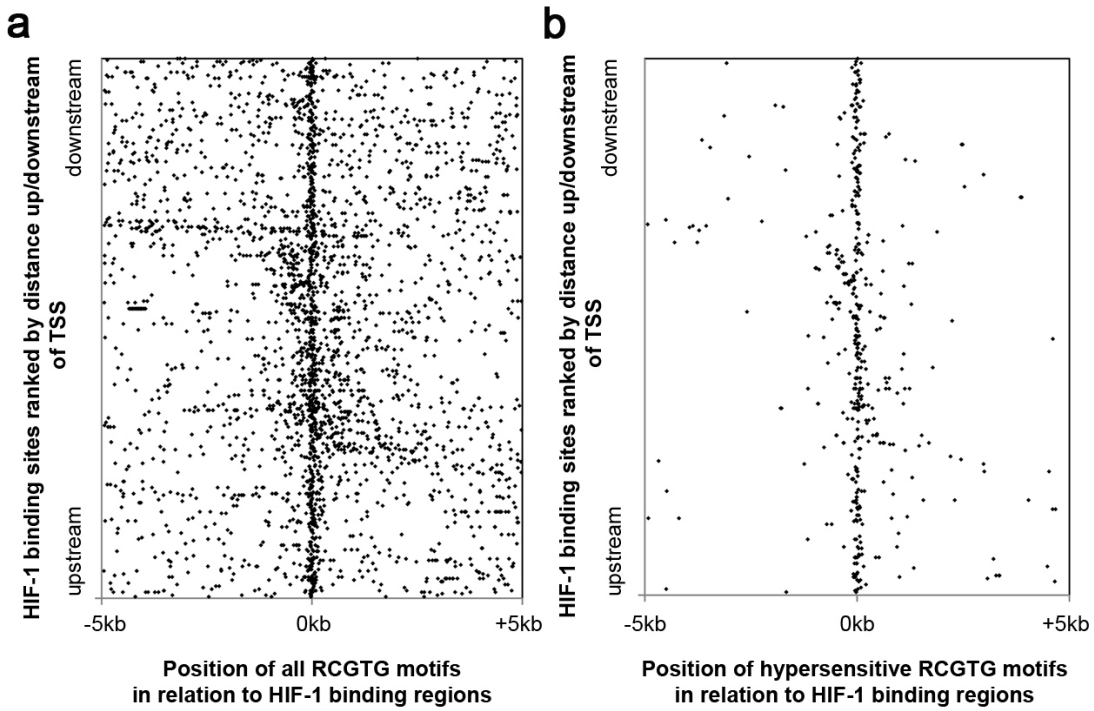


Figure 2.18: Normoxic DNAse1 hypersensitivity predicts hypoxic binding of HIF. High-stringency HIF-1 binding regions were ranked according to distance up/downstream of the nearest TSS (vertical axis). (a) For each HIF-1 binding region, the position of each RCGTG motif, relative to the centre of the HIF-binding region, was plotted (horizontal axis). (b) The same analysis was repeated for the subset of RCGTG motifs that coincided with a normoxic DNAse1 hypersensitivity peak (ENCODE consortium dataset 1).

The data therefore suggests that genome-wide analyses of DNAse1 hypersensitivity may be used to assist in the identification of functionally important RCGTG motifs at specific loci. As an example, the HIF-1α and HIF-2α ChIP-seq data are shown at the *EGLN3* gene locus, together with the positions of all RCGTG motifs and of just the DNAse1 hypersensitive RCGTG motifs (Figure 2.19). Of the multiple RCGTG motifs at or close to the *EGLN3* gene locus, only one coincides with a normoxic DNAse1 hypersensitivity peak. This hypersensitive motif (which corresponds to a functional regulatory sequence defined in previous analyses) binds both HIF-1α and HIF-2α isoforms, whilst none of the insensitive RCGTG motifs in this region binds HIF.

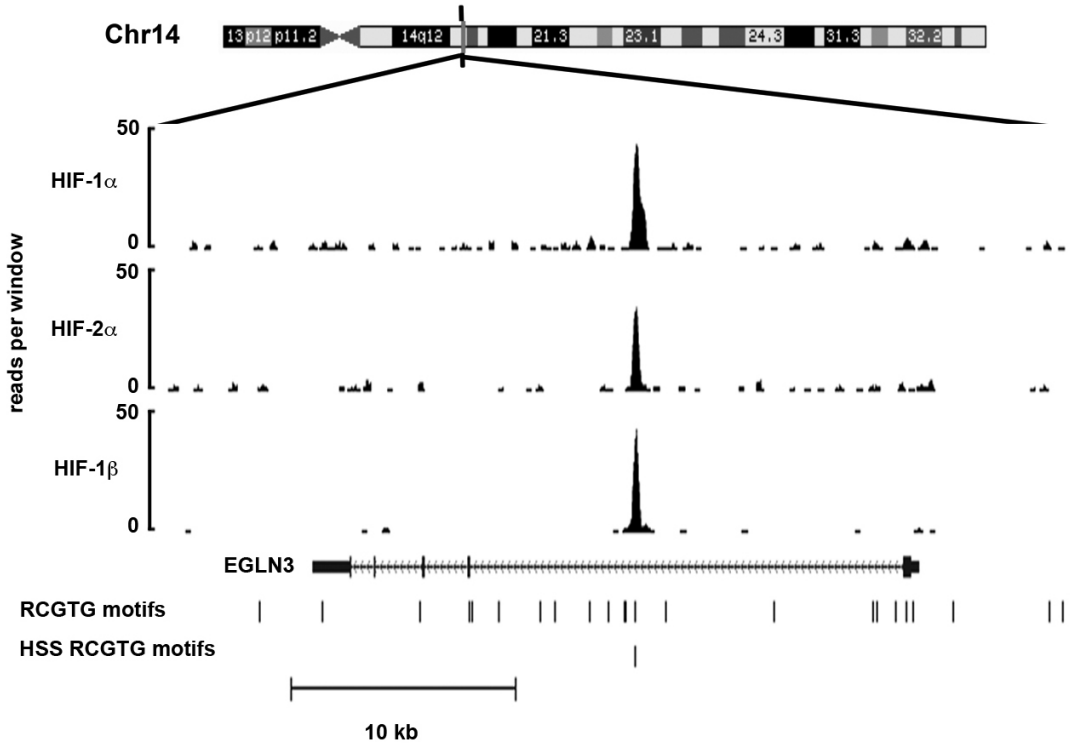


Figure 2.19: Genomic binding of HIF-1α, HIF-2α and HIF-1β subunits at the *EGLN3* gene locus. The number of reads per 100 base pair window and the location of the gene together with the positions of the RCGTG motifs, and DNAse1 hypersensitive RCGTG (HSS RCGTG) motifs are shown for comparison.

Thus, accessibility as indicated by normoxic DNAse1 hypersensitivity is a major factor in predicting HIF-binding to an RCGTG motif. This also indicates that most HIF-binding sites are already accessible for transcription factor binding in normoxia and that "opening up" of sites for/by HIF-binding is a rare event if existent at all. However, despite this, the proportion of hypersensitive RCGTG motifs that bind HIF-α remains low (1.1% of hypersensitive RCGTGs bind HIF-1α and 1.0% bind HIF-2α), indicating that other epigenetic mechanisms also likely contribute to the definition of a HIF-binding site.

2.2.9 Correlation of HIF-binding sites with CpG islands

The previous analyses showed that promoter-distant HIF-binding is associated with enrichment of RCGTGs and DNAse1 hypersensitivity. A similar comparison was undertaken to test whether a common feature of promoters - CpG islands - was enriched

at these sites. This could indicate whether the remote sites are associated with cryptic promoters or represent enhancer sites. When HIF-binding sites were ranked according to their distance to the next TSS a clear association of CpG islands with promoter-proximal sites was noted (Figure 2.20). However, no enrichment of CpG islands could be detected at the remote binding sites indicating that these sites which are associated with RCGTGs and DNAse1 hypersensitivity are regulatory regions outside of CpG islands containing promoters. The effect was more obvious for HIF-2 binding than for HIF-1 binding for which a weak association with CpG islands was seen outside known promoters.

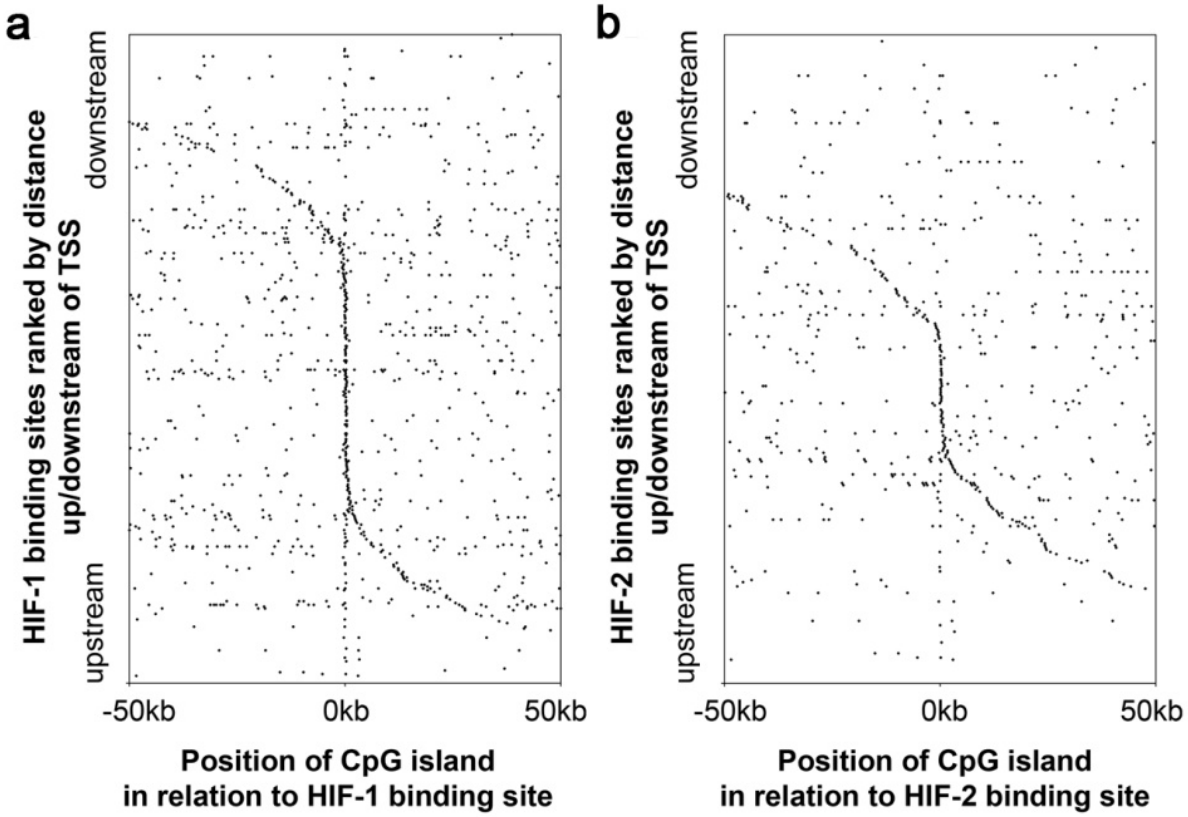


Figure 2.20: CpG islands are associated with promoter-proximal HIF-binding. High-stringency HIF-1 (a) and HIF-2 (b) binding regions were ranked according to distance up/downstream of the nearest TSS (vertical axis). (a) For each HIF-1 binding region, the position of an annotated CpG island (UCSC data base), relative to the centre of the HIF-binding region, was plotted (horizontal axis). (b) The same analysis was repeated for HIF-2 binding sites.

2.2.10 Normoxic HIF-1β binding

HIF-1β has been described to heterodimerise with a variety of binding partners, most notably the AhR or SIM transcription factors (Swanson 2002). AhR like the HIF-α isoforms is a transcription factor induced by environmental stimuli. AhR is expressed in MCF-7 cells and it has been shown to compete with HIF-α for HIF-1β under certain conditions (Chan, Yao et al. 1999). In the current work, induction of AhR was not performed, but it is conceivable that HIF-1β is binding DNA under normoxic conditions together with another transcription factor or as a homodimer (Sogawa, Nakano et al. 1995; Swanson and Yang 1999). To test this, normoxic HIF-1β peaks were examined and no significant overlap was seen, amongst the highest peaks, with either the hypoxic HIF-1β binding sites, normoxic hypersensitive sites or transcriptional start sites (Figure 2.21). This lack of overlap with the hypoxic HIF-1β sites indicates an absence of constitutive HIF-1β binding in these cells. Furthermore, the lack of overlap of the normoxic HIF-1β binding sites with a marker for open chromatin or promoter regions points to no functional relevant HIF-1β binding under these conditions. Amplification of "background" DNA in these samples in the library preparation step is most likely to account for the normoxic HIF-1β signals. Taken together, in MCF-7 cells under these conditions HIF-1β only binds in hypoxia (DMOG) and only at loci where HIF-α subunits are also present.

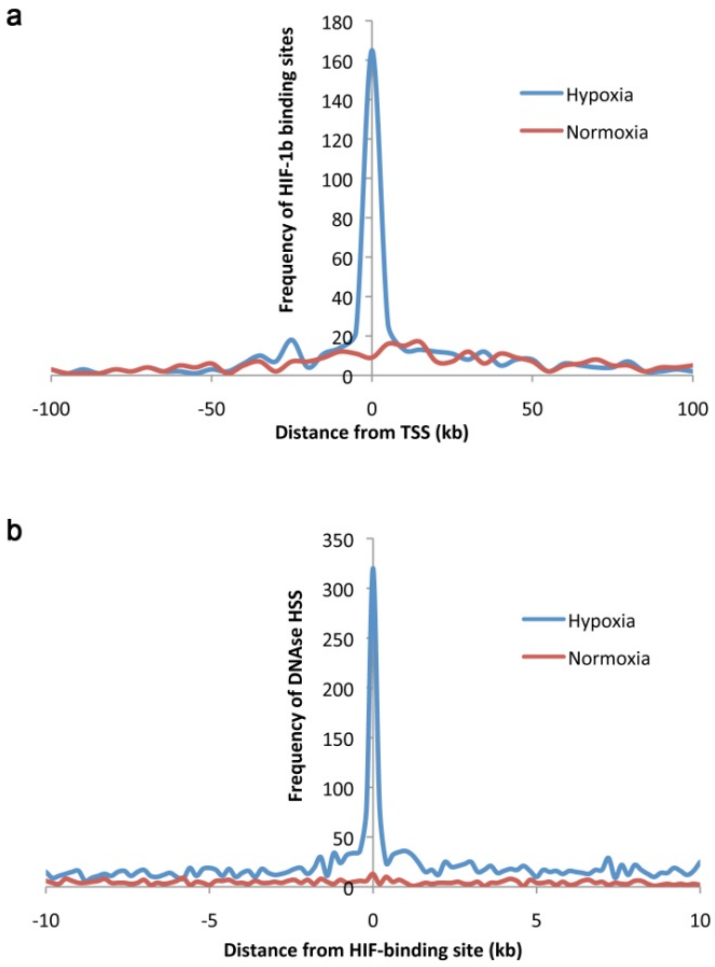


Figure 2.21: (a) Overlap of hypoxic and normoxic HIF-1β peaks with transcriptional start sites (TSS). Frequency of HIF-1β binding sites with respect to the TSS of the next nearest gene is shown for hypoxia (blue) and normoxia (red). (b) Frequency of normoxic hypersensitive sites at hypoxic and normoxic HIF-1β peaks is shown. Note that the lack of overlap with common features of transcription factor binding sites (TSS, HSS) indicates the absence of normoxic HIF-1β binding in MCF-7 cells.

2.2.11 HIF-binding and gene regulation

Since HIF-binding sites were identified in all regions of the genome, including many at great distances from gene promoters, the relationships between the position of HIF-binding sites and effects on gene expression were examined next.

To first test the relationships of HIF-binding with gene regulation under hypoxic conditions (0.1% O_2) or in cells treated with DMOG (1 mM) to stabilise HIF the transcriptome of MCF-7 cells was profiled using mRNA-sequencing (mRNA-seq). mRNA

from cells was isolated and forwarded to the sequencing facility at the Wellcome Trust Centre. After poly-A selection, fragmentation, reverse transcription and size selection, cDNA was sequenced and reads mapped to the human genome (hg18). Initial inspection of individual gene loci suggested valid results. As outlined above EGLN3 expression is highly inducible by hypoxia and DMOG via the HIF-pathway. Figure 2.22 shows mRNA-seq data at the *EGLN3* locus and visualises low levels of EGLN3 expression in normoxia with a strong increase of exonic reads after HIF-stabilisation.

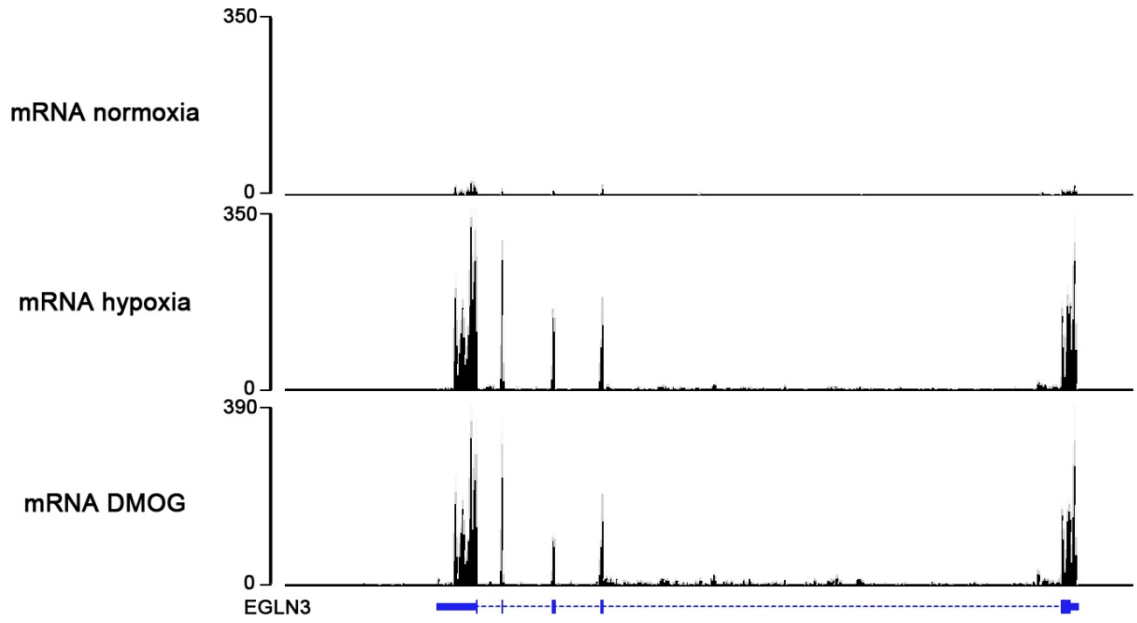


Figure 2.22: mRNA-seq tracks showing read density at the *EGLN3* locus in normoxia and hypoxia (0.1% O_2) and under DMOG 1 mM treatment in MCF-7 cells. Note the good overlap of exon structure with read distribution and the strong increase of exonic reads under hypoxia and DMOG.

To further test validity of the mRNA-seq data a comparison with genome-wide expression arrays was undertaken (Elvidge, Glenny et al. 2006). In gene set enrichment analyses (GSEA, (http://www.broadinstitute.org/gsea/)) genes upregulated by hypoxia in the microarray data set clustered with genes upregulated by hypoxia in the mRNA-seq data set (Figure 2.23). The same enrichment was seen for genes downregulated by hypoxia in

the chip analysis in genes downregulated in the mRNA-seq data set. These findings confirm that the mRNA-seq data is a good correlate of hypoxic gene regulation.

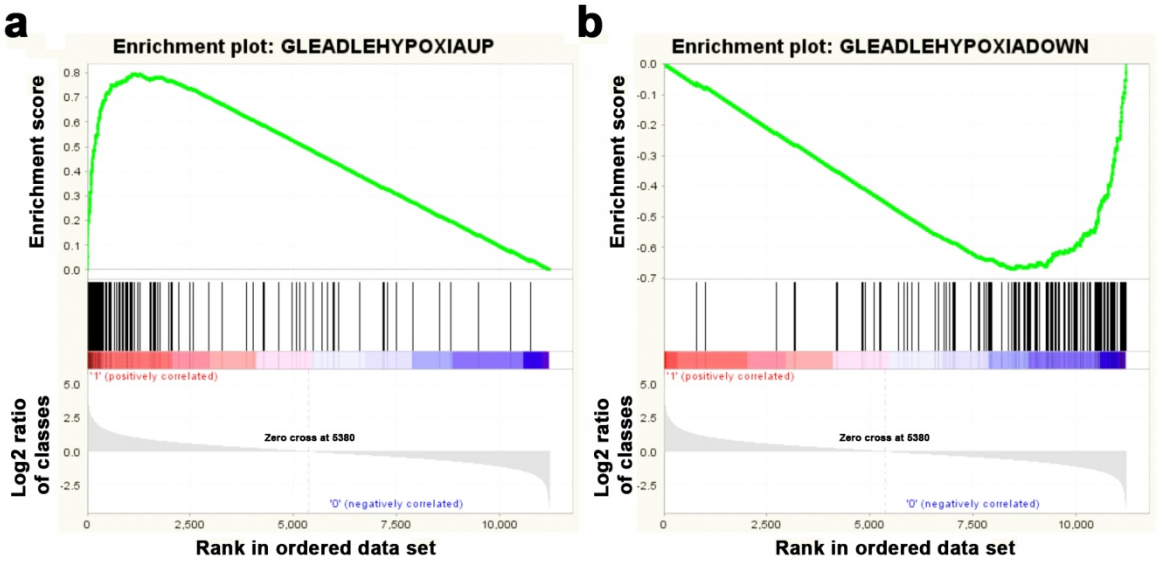


Figure 2.23: Gene Set Enrichment Analysis comparing mRNA-seq with results from microarray studies (Elvidge, Glenny et al. 2006). (a) Genes from the mRNA-seq data set were ranked according to fold regulation by hypoxia (x-axis, red indicates upregulation, blue indicates downregulation) and a set of genes upregulated by hypoxia in the microarray study was tested for enrichment. The cumulative enrichment score together with the position of the gene amongst the genes ranked by fold regulation in the mRNA-seq data set is shown. Upregulated genes in the microarray set cluster with upregulated genes in the mRNA data set. (b) Similar analysis using a set of downregulated genes in the microarray study.

To examine the relationships of HIF-binding with hypoxic gene regulation further gene set enrichment analyses were performed by intersecting the above identified high stringency HIF-1 and HIF-2 binding sites with a series of expression data sets. A strong enrichment of HIF-binding sites (HIF-1 and HIF-2) amongst genes upregulated by hypoxia or DMOG could be detected in the mRNA-seq data set (Figure 2.24). This is in concordance with previous results from this laboratory (Mole, Blancher et al. 2009), and with published data (Xia, Lemieux et al. 2009) which show that HIF-binding strongly correlates with transcriptional activation rather than repression. Taken together, these analyses confirm the validity of the genome-wide mRNA-seq data sets and implicate that HIF-binding is associated with activation of gene transcription. However, the previous findings from ChIP-seq

experiments revealed a high number of HIF-binding sites remote from the next nearest annotated gene. This raises the question whether HIF-binding remote from genes is functional or not. The fact that HIF-binding remote from promoters is occurring in regions of open chromatin creates the possibility that HIF-binding has long-range effects on gene transcription presumably via binding to functional enhancers.

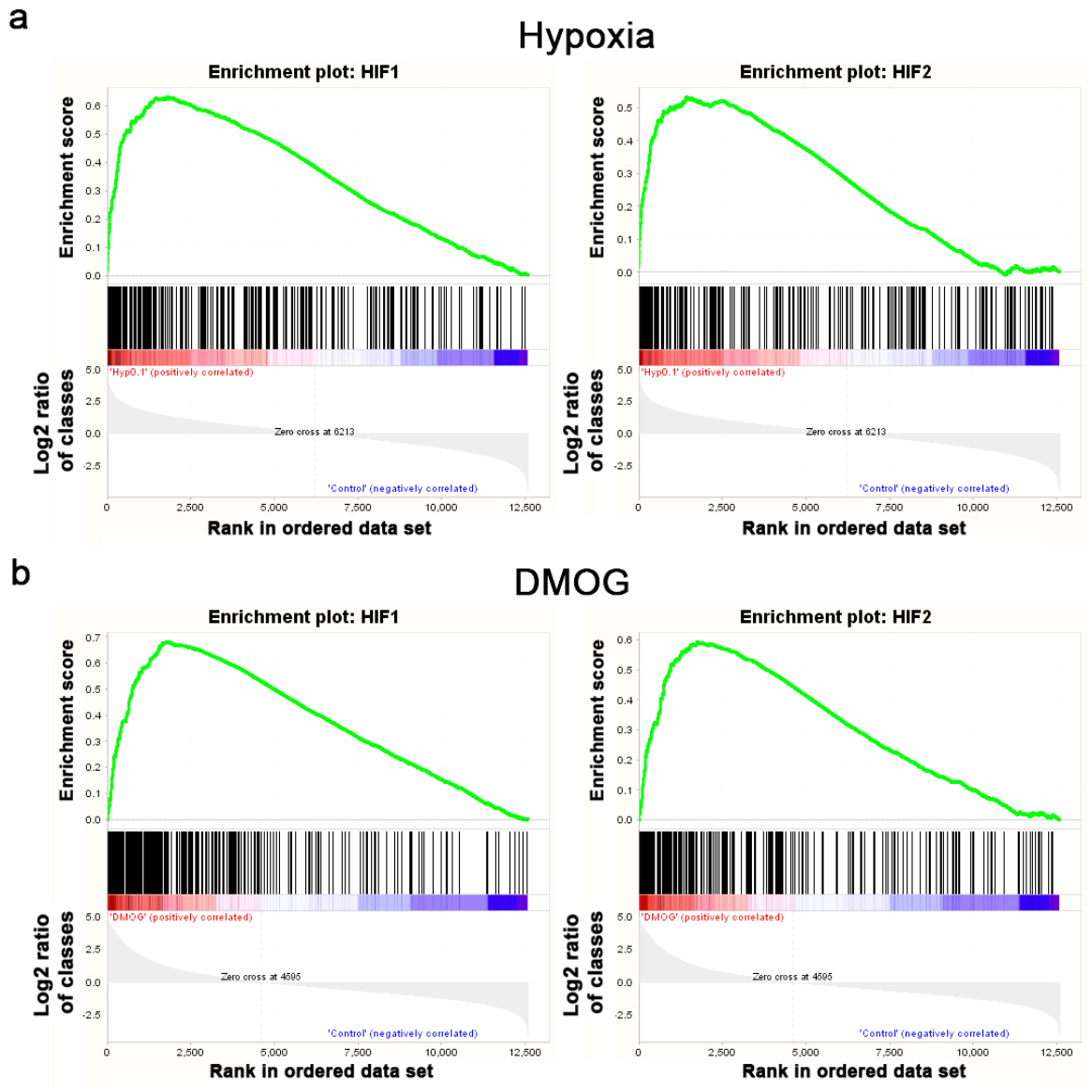


Figure 2.24: Gene Set Enrichment Analysis of HIF-1 and HIF-2 binding gene loci. mRNA-seq data (FPKM values, fragments per kilobase of exon model per million mapped fragments (Mortazavi, Williams et al. 2008) filtered for present calls and ranked according to fold regulation by hypoxia (0.1% O_2) or DMOG (1 mM) were used, and GSEA analysis was performed. (a) The cumulative enrichment score together with the position of the HIF-binding gene loci amongst the genes ranked by fold regulation by hypoxia are plotted for all high-stringency HIF-1 binding gene loci (left) and all high-stringency HIF-2 binding gene loci (right). The red region indicates genes upregulated and the blue region genes downregulated by hypoxia. (b) Similar analysis using mRNA-seq data for DMOG treated cells.

To first test whether HIF-binding is directly influencing gene transcription irrespective of the distance to the next nearest gene a series of comparisons was performed. Previously Elvidge et al published expression array data from MCF-7 cells incubated in 1% O_2 and treated with control siRNA or combined HIF-1α and HIF-2α siRNAs (Elvidge, Glenny et al. 2006). This presents a more accurate correlate of HIF-dependent gene regulation than hypoxia or DMOG alone and was used to define the fold-regulation by HIF. Using these data sets the high stringency HIF-binding gene loci were tested for enrichment amongst HIF-regulated genes using the GSEA platform. HIF-regulated gene loci were identified based on the closest gene (TSS) to each binding site. As expected gene loci binding HIF-1α and HIF-2α strongly clustered amongst the genes that were most highly upregulated by HIF. In contrast no enrichment of HIF-binding gene loci was observed amongst the genes downregulated by HIF (Figure 2.25).

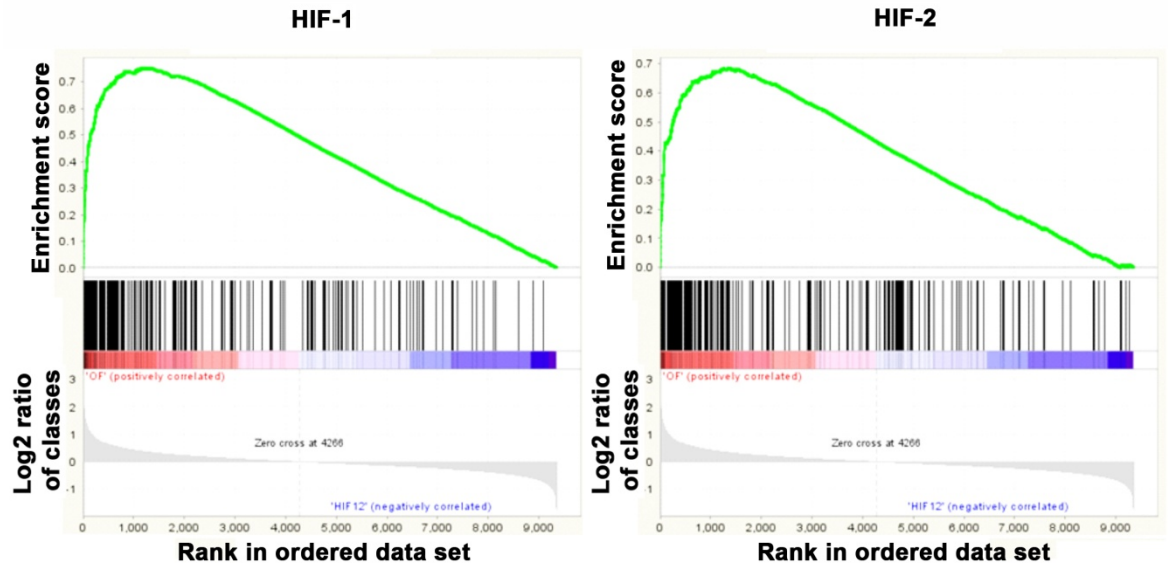


Figure 2.25: Gene Set Enrichment Analysis of HIF-1 and HIF-2 binding gene loci. Expression array data from HIF-1/2 knockdown experiments (Elvidge, Glenny et al. 2006) were used, and GSEA analysis was performed. The cumulative enrichment score together with the position of the HIF-binding gene loci amongst the genes ranked by fold regulation are plotted for all high-stringency HIF-1 binding gene loci (left) and all high-stringency HIF-2 binding gene loci (right). The log2 fold regulation by HIF-α is illustrated at the bottom of each panel. The red region indicates genes upregulated and the blue region genes downregulated by HIF as determined by combined HIF-1/2 siRNA knockdown.

To examine the relationship between HIF-binding remote from promoters and transcriptional regulation, HIF-1α binding gene loci were next stratified according to the distance between the TSS and the nearest HIF-1α binding site and further tests were run for association with upregulation by HIF, using GSEA. These analyses revealed that gene loci that bound HIF-1α were significantly associated with upregulation of the nearest transcript by HIF, irrespective of the distance between the HIF-1α binding site and the TSS (Figure 2.26). The association was robust even for loci where the HIF-1α binding site was greater than 100 kb from the nearest TSS, indicating that HIF-1α DNA-binding can exert effects on gene regulation over distances of greater than 100 kb. Comparable results were seen when the analysis was repeated for HIF-2α (Figure 2.26, right panel).

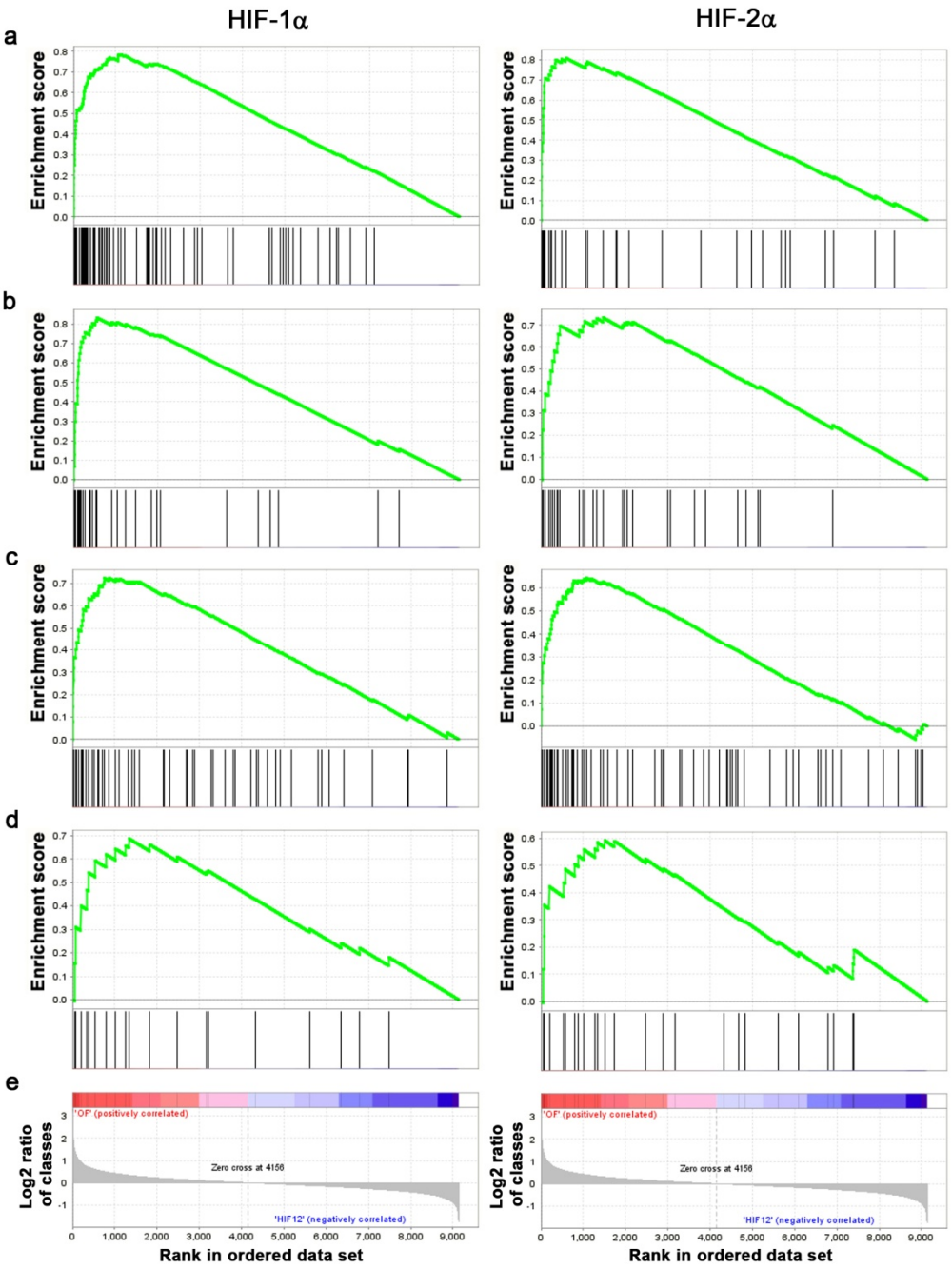


Figure 2.26 Gene Set Enrichment Analysis of HIF-2 binding gene loci stratified by distance between binding site and TSS. Expression array data as for Figure 2.25 were used. HIF-1 and HIF-2 binding gene loci were stratified according to the distance to the nearest TSS into 4 sub groups and GSEA analysis performed as before. The cumulative enrichment score (ES), together with the position of the HIF-binding gene loci amongst the genes ranked by fold regulation are plotted for (a) gene loci with the closest HIF-binding within 1 kb of the TSS, (b) between 1 kb and 10 kb of the TSS, (c) between 10 kb and 100 kb of the TSS and (d) greater than 100 kb from the TSS. (e) The log2 fold regulation by HIF-α is illustrated.

To pursue and validate these findings further, I performed additional analyses on two of the most distant HIF-binding sites. HIF DNA-binding at both the *ARRDC3* and *ERRFI1* loci is

located more than 100 kb distant from the promoter, although in each case these are the nearest documented genes (Figure 2.27a and b) and both have been reported to be regulated by HIF (Elvidge, Glenny et al. 2006). Two significant HIF-binding peaks were present at the *ERRFI1* locus and one at the *ARRDC3* locus. Each peak confirmed in ChIP-qPCR analysis with high levels of enrichment (Figure 2.27c). Furthermore, analysis of mRNA levels by qPCR, confirmed hypoxia dependent regulation of these transcripts (Figure 2.27d). ChIP-qPCR analysis of these sites using an antibody directed against RNA polymerase 2 demonstrated that the interaction of these loci with RNA polymerase 2 was increased following 16 hours at 0.5% hypoxia (Figure 2.27e) thus indicating hypoxia-inducible interaction between these binding sites and the general transcriptional machinery. Finally, fusion of DNA-sequences spanning these binding sites to a heterologous luciferase reporter gene conferred hypoxic regulation on luciferase activity, confirming the ability of these sequences to mediate hypoxia inducible gene expression (Figure 2.27f).

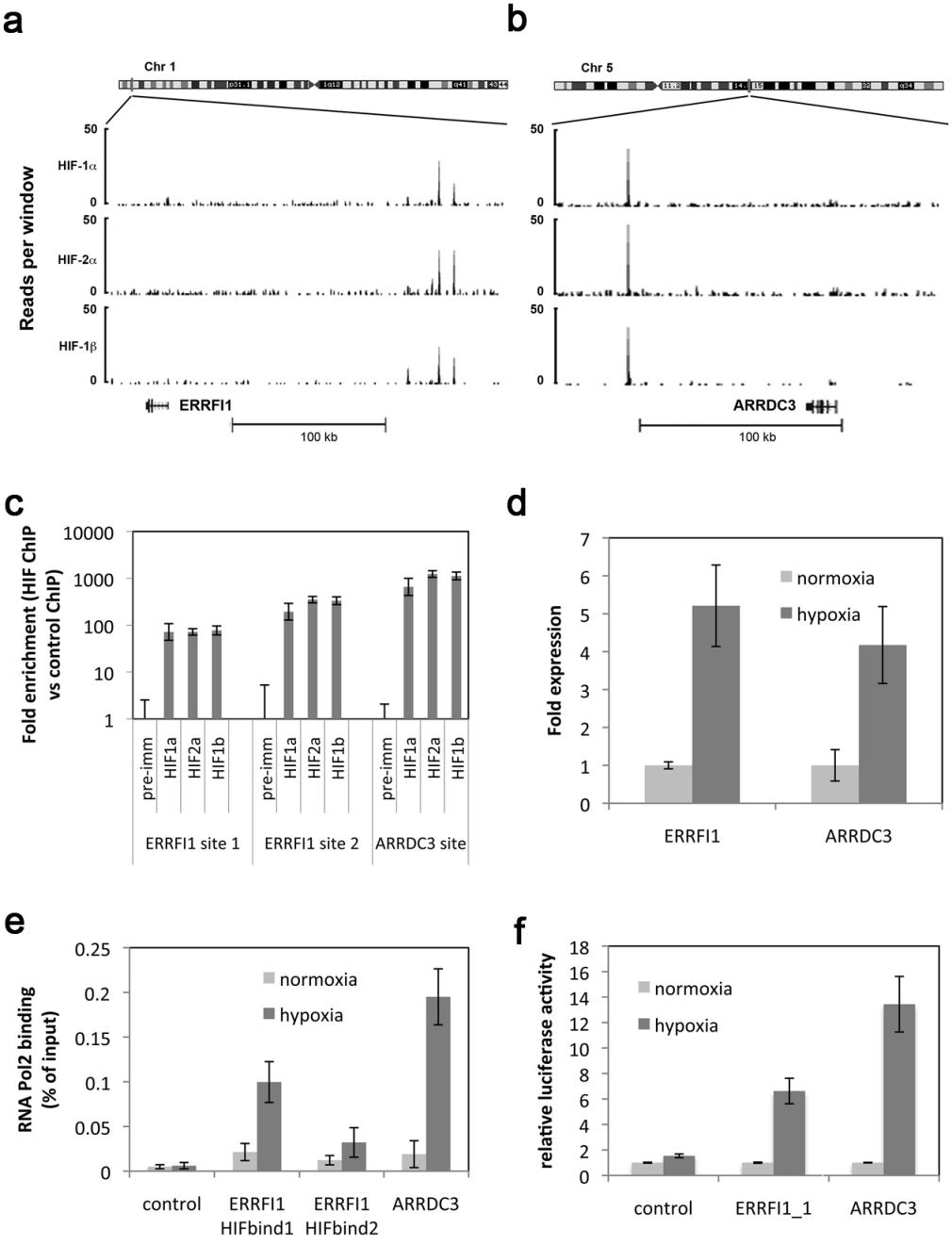


Figure 2.27: HIF-binding regulates gene expression over long genomic intervals. Genomic binding of HIF-1α, HIF-2α and HIF-1β subunits at the (a) *ERRFI1* and (b) *ARRDC3* gene loci. The number of reads per 100 base pair window and the location of each gene are shown. (c) ChIP-qPCR confirmation of the ChIP-seq defined binding peaks. (d) Expression-qPCR confirmation of gene regulation in response to 16 hours at 0.5% hypoxia. (e) ChIP-qPCR demonstrating RNA pol2 recruitment to the *ERRFI1* and *ARRDC3* HIF-binding sites in normoxia and after 16 hours at 0.5% hypoxia. (f) Luciferase reporter assay showing activity of a heterologous luciferase reporter gene fused to sequences spanning each HIF-binding site (278 bp for ERRFI1_1 and 276 bp for ARRDC3) in normoxia and after 16 hours at 0.5% hypoxia.

Given this evidence that binding of HIF to DNA can exert functional effects on gene regulation over distances in excess of 100 kb, and in the light of the observation that a

large proportion of the genes nearest to a HIF DNA-binding site are not themselves regulated by HIF, the association between the binding of HIF and regulation of next-nearest transcript was examined next. To do this, HIF DNA-binding sites at which the nearest transcribed locus was regulated by HIF were first excluded from the analyses. For the remaining HIF DNA-binding sites, the next nearest transcribed locus was then used in analyses of association using the GSEA software. These next nearest HIF-binding gene loci also showed significant association with transcripts regulated by HIF indicating that HIF that is bound to DNA commonly 'skips' the nearest gene to regulate another more remote promoter (Figure 2.28).

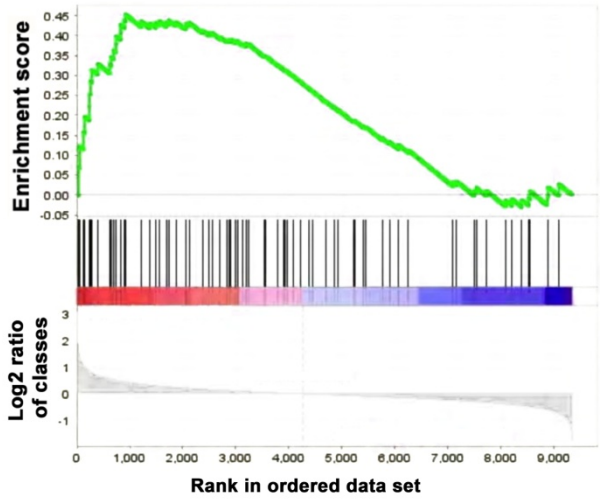


Figure 2.28: Gene Set Enrichment Analysis of HIF-1 binding loci from which the nearest gene is not regulated by HIF. The enrichment analysis indicates that HIF-binding is still enriched among regulated genes even if the next nearest gene is not regulated.

HIF-binding loci were then grouped according to whether the HIF-1α binding was in the intron, exon, 5' UTR, 3' UTR, upstream from the TSS or downstream from the TES. Each group of HIF-binding gene loci was then examined for association with transcriptional regulation by HIF at that locus, in a similar way using GSEA. The set of genes that bound HIF in the 3' UTR was too small for meaningful analysis. All of the remaining groups of genes showed significant association (Figure 2.29).

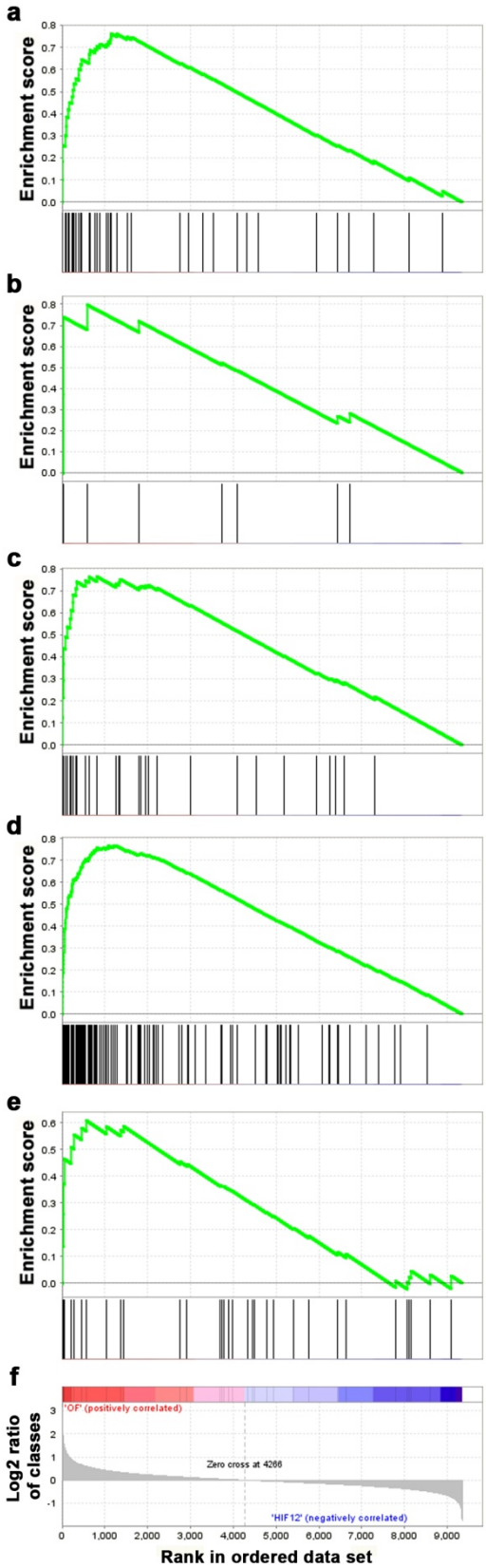


Figure 2.29: Gene Set Enrichment Analysis Expression array data as for Figure 2.25 were used. HIF-1 binding gene loci were stratified according to the position within the gene and GSEA analysis performed as before. The cumulative enrichment score (ES), together with the position of the HIF-binding gene loci amongst the genes ranked by fold regulation are plotted for binding sites situated in the (a) intron, (b) exon, (c) 5' UTR (d) upstream of the TSS and (e) downstream of the TES. (f) The log2 fold regulation by HIF-α is illustrated.

Finally, the association between the amplitude of upregulation of a transcript by HIF and the distance of the nearest HIF-binding site to the promoter was examined; this analysis revealed that the amplitude of transcriptional upregulation by HIF bore no relationship to the position of the nearest HIF-binding site to the promoter (Figure 2.30). Taken together, these results indicate that, at a genome-wide level, HIF-binding loci appear to be functionally similar i.e. they exert qualitatively and quantitatively similar effects on hypoxia inducible gene expression over large genomic intervals irrespective of position with respect to the regulated gene promoter.

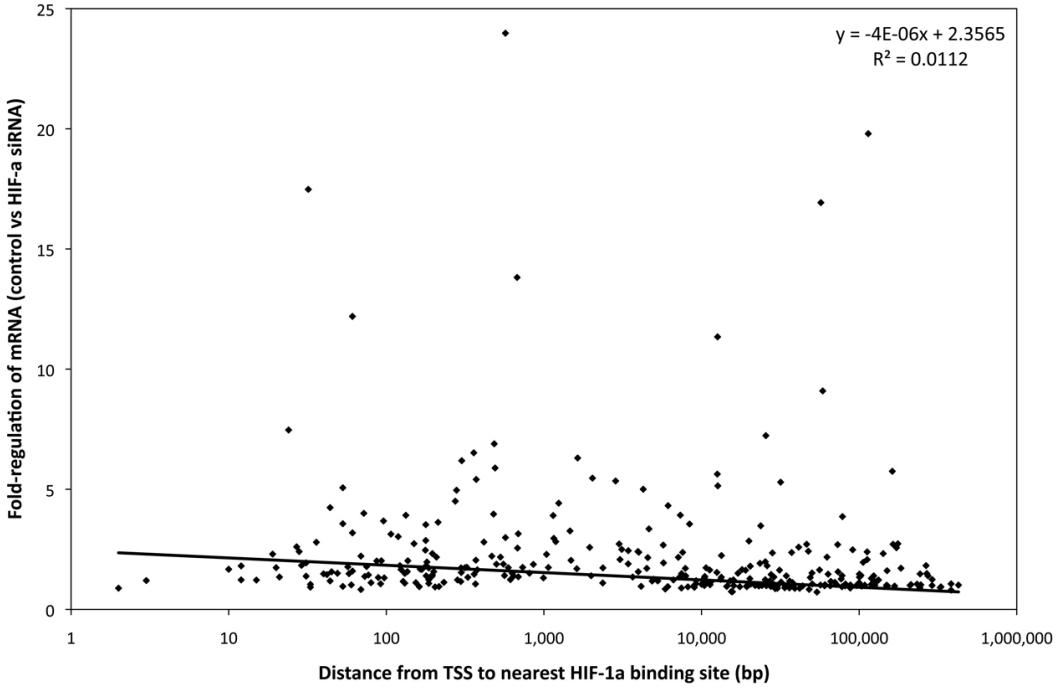


Figure 2.30: For each HIF-1 binding site, the fold-regulation by HIF of the nearest gene was plotted against the distance between the binding site and the TSS.

2.3 Discussion

This chapter describes the use of chromatin immunoprecipitation coupled to next-generation high-throughput sequencing to define HIF-α and HIF-1β binding sites across the whole genome in MCF-7 cells. The technique enables true genome-wide coverage and more precise localization of HIF-binding sites than has previously been possible, for example by using tiled microarrays. The use of MCF-7 cells has allowed comparison with genome-wide studies of hypoxia-inducible gene expression by RNA-seq or microarrays (Elvidge, Glenny et al. 2006), and high-resolution mapping of normoxic DNAse1 hypersensitivity sites in the same cell line.

2.3.1 HIF-α and HIF-β binding in MCF-7 cells

The high degree of concordance between HIF-α binding in cells exposed to DMOG, and hypoxic HIF-1β binding indicates that both stimuli produce comparable genome-wide patterns of HIF DNA-binding which is in line with the high degree of overlap between the transcriptional responses to DMOG and hypoxia (Elvidge, Glenny et al. 2006; Warnecke, Weidemann et al. 2008). Furthermore, the strong concordance observed between HIF-α and HIF-1β binding implies that either HIF-α can only bind DNA as a heterodimer with HIF-1β or that if HIF-α is able to bind DNA with an alternative partner (e.g. as part of an Sp-1 binding complex) that it does so at the same sites as the αβ-heterodimer, with much lower affinity/occupancy, or at too few sites to be identified in a genome-wide analysis. Conversely, it also indicates that in these cells under these conditions the large majority of HIF-1β is bound as a HIF-α/β heterodimer. Earlier reports suggest that HIF-1β can bind DNA as a homodimer (Sogawa, Nakano et al. 1995; Swanson and Yang 1999) and more recently a role for ARNT in estrogen receptor pathways has been discovered (Swedenborg

and Pongratz 2010). In the present study there is no evidence for widespread HIF-1β DNA-binding independent of HIF-α under basal conditions. Binding peaks identified under normoxic conditions in which HIF-α is absent did not correlate with common features of transcription factor binding sites such as TSS or DNAse1 HSS. However, these findings do not exclude DNA-binding of HIF-1β with other dimerisation partners in different circumstances such as induction of the AhR or the estrogen receptor system and a potential competition of HIF-α subunits with other dimerisation partners for HIF-1β. Neither possibility was tested in the current study. Taken together, both HIF-α and HIF-β binding is inducible, occurs on the same sites most probably as a heterodimer and there is no evidence for heterodimerisation with other binding partners or homodimerisation for HIF-1β.

2.3.2 Definition of HIF-binding sites

Using highly stringent criteria, more than 500 HIF-binding sites were identified across the genome. Many of these sites were outside of regions covered by promoter arrays and were remote from core promoters. Indeed HIF-binding sites were observed many hundreds of kilo bases distant from any documented promoter. Similar enrichment was observed for both the core HIF-binding RCGTG motif and for DNAse1 hypersensitive sites amongst these promoter-distant binding sites, when compared to promoter-proximal sites. Together with independent experimental validation by qPCR, these results indicate that the identification of promoter-distant sites was accurate, and likely to be functionally relevant. The number of high stringency binding sites defined in this study is comparable to numbers found for HIF-1 and HIF-2 in ChIP-on-chip studies in the same cell line (Mole, Blancher et al. 2009) and for HIF-1 in genome-wide analyses in other cell lines (Xia,

Lemieux et al. 2009; Tanimoto, Tsuchihara et al. 2010). It is also similar to the number of sites identified for other transcription factors such as p53 in HCT116 colorectal cancer cells (Wei, Wu et al. 2006) or FOXP3 in mouse T-cells (Zheng, Josefowicz et al. 2007). In contrast other transcription factors appear to have significantly higher binding frequencies. For example, many thousands of binding sites have been identified for the estrogen receptors (Carroll, Meyer et al. 2006; Fullwood, Liu et al. 2009; Welboren, van Driel et al. 2009; Grober, Mutarelli et al. 2011) M Nature 2009), E2F1, Myc (Bieda, Xu et al. 2006) and Stat1 (Heintzman, Hon et al. 2009). All of these studies described transcription factor binding at promoter-distal sites as common indicating that long-range enhancer promoter interaction is a widespread mode of gene regulation.

The relative modest number of binding sites implicates that the HIF-pathway is regulating a well defined selection of genes directly and raises the possibility that clusters of HIF activated genes exist in the genome. However, nearest neighbour analysis of HIF-binding and HIF-regulated genes revealed a distribution of distances comparable to randomly chosen control loci. This lack of clustering of both HIF-binding and particularly HIF-regulated gene loci provides evidence against widespread operon-like regulation of genes or genomic clustering of HIF transcriptional targets. However, the possibility that distant chromatin elements cluster in three-dimensional space to form hypoxia inducible "transcription factories" cannot be excluded.

2.3.3 Genetic and epigenetic environment at HIF-binding sites

The improved accuracy and unbiased ascertainment of HIF-binding across the genome also permitted significant refinements to be made to the HIF consensus (Wenger, Stiehl et al. 2005). These studies revealed preferences at a number of sites outside the RCGTG consensus including the relative exclusion of the E-box motif CACGTG, suggesting that

competitive binding by E-box proteins may exclude HIF from such sites even in hypoxic cells.

Notably, it was not possible to distinguish HIF-1 from HIF-2 binding motif preferences even for residues outside the RCGTG consensus, despite the fact that HIF-2 binding sites were differently distributed from HIF-1 binding sites, being substantially more prevalent at sites remote from promoters and targeting different pathways. This would suggest that such motifs if they exist are either too pleiotropic or too distant to detect in motif analysis of HIF-binding regions or that non-sequence defined, "epigenetic" chromatin markers contribute to the differential binding of the two isoforms.

Compared to some other transcription factors that have been analysed at the pan-genomic level, the number of high stringency HIF-binding sites defined across the genome was relatively modest (Fullwood, Liu et al. 2009). This contrasts with relatively promiscuous primary sequence preferences for HIF-binding, again suggesting that epigenetic processes must determine the binding of HIF to DNA, and may be particularly important in constraining the transcriptional response to hypoxia. Since HIF-α proteins are reduced by proteolysis to very low levels in normoxia, our finding that the majority of HIF-binding sites coincided with normoxic DNAse1 hypersensitivity sites was somewhat surprising, and argues against widespread opening of chromatin conformation by HIF-binding or hypoxia itself. This finding is, however, consistent with observations that the HIF-binding site at the *erythropoietin* (*EPO*) locus corresponded to a normoxic DNAse1 hypersensitive region (Semenza, Nejfelt et al. 1991), that chromatin accessibility can pre-determine glucocorticoid receptor binding patterns (John, Sabo et al. 2011), and that the pattern of HIF-binding at promoters correlates with markers of normoxic transcriptional activity at these loci (Xia and Kung 2009). In addition, a recent study using ChIP-seq to

identify HIF-1 binding sites in two different cell lines and Nucleosome-seq to identify regions of open chromatin confirms that an open chromatin conformation pre-exists at HIF-binding sites under normoxic conditions (Tanimoto, Tsuchihara et al. 2010). However, in the current study many DNAse1 hypersensitive HRE motifs both close to and distant from actively transcribed genes did not bind HIF. This suggests that additional epigenetic factors determining the distribution of these enhancer sites in different cellular contexts are likely to have major influence on the transcriptional response to hypoxia.

2.3.4 HIF-binding and gene regulation

Ingenuity Pathways Analysis demonstrated that many canonical pathways are targeted by HIF with varying bias towards either HIF-1 or HIF-2 binding. Consistent with previously published work, pathways involved in carbohydrate metabolism are principally targeted by HIF-1, whilst genes with a role in Oct-4 regulated stem cell pluripotency predominantly bind HIF-2 (Covello, Kehler et al. 2006). Interestingly, pathway-specific association of particular HIF isoforms was observed even though siRNA studies have suggested that in MCF7 cells HIF-2 appears to have rather little transcriptional activity, suggesting that isoform-specific patterns of HIF-binding might extend across different cells and is consistent with post-DNA-binding mechanisms limiting transcriptional activity (Hu, Sataur et al. 2007; Lau, Tian et al. 2007).

Correlation of HIF-binding data with functional studies of the HIF transcriptional response using GSEA revealed the existence of robust statistical associations with patterns of hypoxia inducible gene expression even over genomic distances in excess of 100 kb. Indeed these studies indicated that for HIF-1 binding, the effects on hypoxia-inducible gene expression were qualitatively and quantitatively similar, irrespective of the position of HIF-1 binding at the gene locus, irrespective of distance at least up to 100 kb, and

irrespective of the presence of an intervening gene that was not responsive to the HIF pathway. This is compatible with a model in which most or all of the promoter-distant HIF-binding sites are hypoxia-inducible enhancers that connect with transcriptional complexes at the promoters of target genes by DNA-looping over potentially very large distances, and thus exert similar effects on the transcriptional activity of these complexes irrespective of physical distance along the chromosome.

The current ChIP-seq analysis confirms and extends observations on the architecture of the hypoxia inducible response that have been derived from ChIP-on-chip analyses using tiled promoter arrays (Mole, Blancher et al. 2009; Xia and Kung 2009; Xia, Lemieux et al. 2009). In particular, even when promoter-distant sites are included in the analyses there is no evidence of an association with the loci of transcripts that were negatively regulated by HIF. This confirms that downregulation of transcript abundance by the HIF pathways is very largely, or even completely, indirect with respect to the binding of HIF to DNA and that HIF is a transcriptional activator. This finding is in contrast to reports for other transcription factors such as p53 (Wei, Wu et al. 2006) or the estrogen receptor (Carroll, Meyer et al. 2006) in which similar numbers of transcripts are directly up or downregulated, respectively.

3 CHAPTER 3: Genome-wide analysis of histone 3 lysine 4 methylation and RNA pol2 in MCF-7 cells under normoxic and hypoxic conditions

3.1 Introduction

In the previous chapter the definition of high stringency HIF-binding sites allowed the comparison of these sites with other genome-wide data sets such as DNAse hypersensitivity sites and CpG islands. These comparisons led to the conclusion that DNAse hypersensitivity in normoxia is a strong predictor of HIF-binding sites. However, despite this, the majority of HREs in DNAse hypersensitive sites still do not bind HIF, suggesting that additional epigenetic mechanisms are involved in influencing whether an HRE binds HIF or not. Recently, analyses of genome-wide data sets for common epigenetic markers, such as histone modifications or RNA pol2, has revealed important information about the epigenetic configuration at active and inactive transcriptional regulatory elements in the genome.

Histone proteins are core constituents of the nucleosome which comprises 146 bp of DNA arranged around a histone octamer. Posttranslational and mainly reversible histone modifications both reflect and influence the activity of gene expression and are probably the major reason why certain genes are transcribed in different developmental stages or in particular cells and others are not (Bannister and Kouzarides 2005; Kouzarides 2007; Bartke, Vermeulen et al. 2010). These histone marks are actively involved in the recruitment of other histone modifying enzymes, transcription factors or members of the basal transcriptional machinery. Therefore, the distribution of these modifications along genes is crucial for transcriptional activity.

For example, histone 3 lysine 4 trimethylation (H3K4me3) is observed in the promoter regions of actively transcribed genes (Guenther, Levine et al. 2007), whereas histone 3 lysine 9 trimethylation (H3K9me3) at the promoter region is a hallmark of silenced genes (Barski, Cuddapah et al. 2007). Similarly, the transcriptional elongation marker,

trimethylation of histone 3 lysine 36 (H3K36me3), is present across the gene body of actively transcribing genes whilst the gene bodies of silent genes are marked with trimethylation of histone 3 lysine 79 (H3K79me3). The validation of these findings in genome-wide approaches revealed a remarkable epigenetic landscape in cells, which is tightly associated with the transcriptional status of genes. Furthermore, genome-wide approaches have allowed the examination of specific regulatory elements in the genome. A series of studies have examined common epigenetic features of regulatory regions (mainly focussing on promoters and enhancers) in the genome. These studies revealed that although transcription factors bind to both promoters and enhancers, the methylation status of histone residues differs at promoters compared to enhancers. Importantly, promoters can be distinguished by high levels of both H3K4me3 and H3K4me1, whereas areas enriched for H3K4me1 and depleted for H3K4me3 more often function as enhancers of gene expression (Heintzman, Stuart et al. 2007). Furthermore, the presence of acetylated H3K27 (H3K27ac) also correlates well with the activity of regions as enhancers (Rada-Iglesias, Bajpai et al. 2011). Thus an analysis of epigenetic marks at promoter-distant HIF-binding sites would help to distinguish true enhancers from cryptic promoters potentially regulating novel and as yet unannotated transcripts.

A second reason for determining pan-genomic patterns of histone modification in hypoxia was to examine whether and in what manner they might themselves be altered by hypoxia/HIF. For example, quantification of bulk histone modifications (i.e. not at any specific gene locus) by immunoblot analyses has demonstrated that specific histone residues are modified under hypoxic conditions (Johnson, Denko et al. 2008; Xia, Lemieux et al. 2009; Zhou, Sun et al. 2010). Secondly, some histone modifying enzymes, including histone demethylases (Beyer, Kristensen et al. 2008; Pollard, Loenarz et al. 2008;

Wellmann, Bettkober et al. 2008; Xia, Lemieux et al. 2009), histone methyltransferases (MLL5) and histone acetyltransferases (SAP30) are themselves transcriptionally regulated by HIF (Mole, Blancher et al. 2009). Finally, amongst the former group, several histone demethylases are 2-OG-dependent dioxygenases enzymes (of the same family as the HIF-hydroxylases) and therefore require oxygen for their activity. These findings further point to a crucial interplay between histone modification and hypoxia-signalling.

In recent reports genome-wide mapping of RNA pol2 has shed more light on the distribution of the transcriptional machinery across the genome. Interestingly, significant signals for RNA pol2 could be detected at both promoters and promoter-distal regulatory elements. Similarly to the findings presented in Chapter 2 for the *ARRDC3* and *ERRFI1* HIF-binding sites recruitment of RNA pol2 to enhancers seems to be a general feature of transcriptional regulation (Heintzman, Stuart et al. 2007) and helps to distinguish active from inactive regulatory elements. Whether these signals reflect looping interaction between enhancers and the basal transcriptional machinery at the promoter of target genes or whether this is a genuine recruitment of RNA pol2 to promoter-distant sites is not completely understood - for example, RNA pol2 activity at enhancers also has been shown to lead to non-coding RNA transcription at enhancers (De Santa, Barozzi et al. 2010). A third possibility for RNApol 2 binding distant from annotated promoters could be binding to non annotated promoters of genes encoding for the recently identified large group of non coding RNAs (Guttman, Amit et al. 2009; Rinn and Chang 2012). These non-coding RNAs (ncRNAs) can directly regulate gene expression or transcript stability (Wutz 2011). Current research is focusing on the functional role of these non-coding RNAs. A prominent example is the *XIST* gene which is responsible for inactivation of one of the X chromosomes during development (Brown, Ballabio et al. 1991; Brown, Lafreniere et al.

1991), More recently, non-coding RNAs regulating *HOX* gene expression (Rinn, Kertesz et al. 2007) or global gene silencing in response to the p53 pathway have been described (Huarte, Guttman et al. 2010).

In addition to RNA pol2 ChIP-seq, high-throughput sequencing of mRNA was used to characterise the abundance of transcripts in normoxic or hypoxic cells. This method allows an unbiased approach to detect any mRNA transcribed in the cell, i.e. not dependent on probes covering annotated genes which are used in microarray analysis. Therefore, besides transcript abundance novel transcripts can be detected and their location compared to HIF-binding sites. In contrast to RNA pol2 ChIP-seq which directly measures RNA pol2 at the DNA, quantification of transcripts by mRNA-seq is also dependent on post-transcriptional modifications of mRNA (e.g. mRNA stability) and therefore adds more information to transcriptional changes in hypoxia.

Given the complex interactions between the epigenetic landscape and transcriptional responses to hypoxia, I next performed a combined pan-genomic analysis of histone H3K4me1 and H3K4me3 modification, and RNA-pol2 binding using ChIP-seq, together with a more complete transcript profiling using mRNA-seq in conjunction with the previous analysis of HIF-binding sites. These investigations have identified general patterns of hypoxic response as well as identifying individual loci, which exemplify different responses of gene regulation to the HIF system and will provide the basis for further analysis.

3.2 Results

3.2.1 Changes of histone modifications under hypoxic conditions

Previous studies have described histone marks (e.g. H3K4me3, H3K9me2 and H3K36me3) that are modified under hypoxic conditions on a bulk level. In order to verify these changes in MCF-7 cells and to potentially identify additional histone marks that are changing I performed immunoblot analysis of a selection of histone modifcations under normoxic and hypoxic (0.5% O_2) conditions using specific antibodies against each modification - most antibodies having been used in previous studies for immunoblot and ChIP experiments (Barski, Cuddapah et al. 2007). Band densities from three independent experiments were measured using an imaging system (Gel-doc) and normalised to signals for total H3. A significant ($p < 0.05$, t-test with Bonferroni correction) increase was determined for H3K4me2, H3K4me3, H3K9me2, H3K9me3, H3K36me3 and H4K20me3 (Figure 3.1). This is consistent with previous reports of hypoxic changes in H3K4me2, H3K4me3, H3K9me2 and H3K36me3 described in other cell types (Johnson, Denko et al. 2008; Xia, Lemieux et al. 2009). In MCF-7 cells, changes in histone marks were most prominent for H3K4me2 and H3K4me3. These modifications are markers enriched at promoters of active genes. To test how these bulk changes are corresponding to specific loci ChIP experiments against selected histone marks at promoters of single genes were performed.

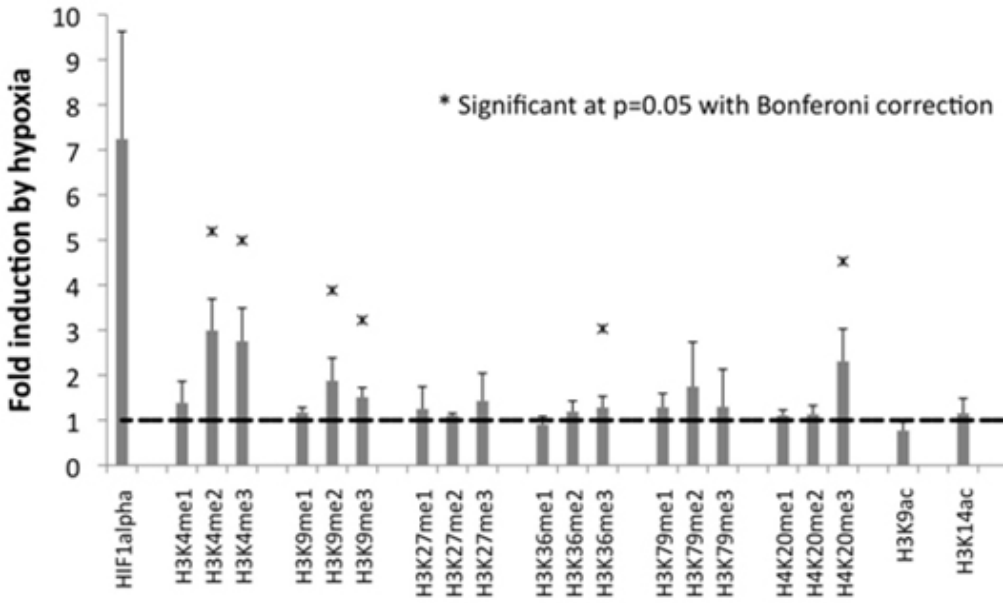


Figure 3.1: Immunoblot analysis of bulk histone marks from MCF-7 cells exposed to normoxia or hypoxia (0.5% O_2 for 16 h). Values are relative band densities of respective histone marks in hypoxia normalised to beta actin levels from three independent experiments. Levels in normoxia were set to 1. Values for HIF-1α are shown for comparison (far left).

3.2.2 Changes of histone modifications and RNA pol2 binding under hypoxic conditions at selected gene loci

As outlined above high levels of histone 3 lysine 4 monomethylation (H3K4me1) and histone 3 lysine 4 monomethylation (H3K4me3) are major features of enhancers or promoters, respectively. This led me to check methylation states of H3K4 (H3K4me1, H3K4me2, H3K4me3) in ChIP experiments at selected promoters under normoxic and hypoxic conditions. Expression of Aldolase C and FAM13A1 has been reported to be upregulated under hypoxic conditions (Elvidge, Glenny et al. 2006) and HIF-binding sites are present in the promoter of both genes in MCF-7 cells (Mole, Blancher et al. 2009; Schodel, Oikonomopoulos et al. 2011). Therefore, ChIP experiments against H3K4me1, H3K4me2, H3K4me3 and RNA pol2 using antibodies which had previously been employed for ChIP-seq experiment (Barski, Cuddapah et al. 2007) were performed to test for

changes at single gene loci. To control between the different conditions ChIP experiments using an antibody against total H3 were included. Recovery of DNA for total H3 was comparable at both gene loci under both, normoxic and hypoxic, conditions. At both promoters, 16 hours incubation in 0.5% hypoxia led to increased capture of promoter DNA by H3K4me3 antibodies (Figure 3.2). This indicates increased trimethylation of H3K4 at these gene loci in hypoxia and in the light of the existing literature increased transcriptional activity, which would be in concordance with increased mRNA levels (Elvidge, Glenny et al. 2006) and an increased recruitment of RNA pol2 to this sites (Figure 3.2). These data suggest that changes in histone modifications are concurrent with gene transcription under hypoxic conditions.

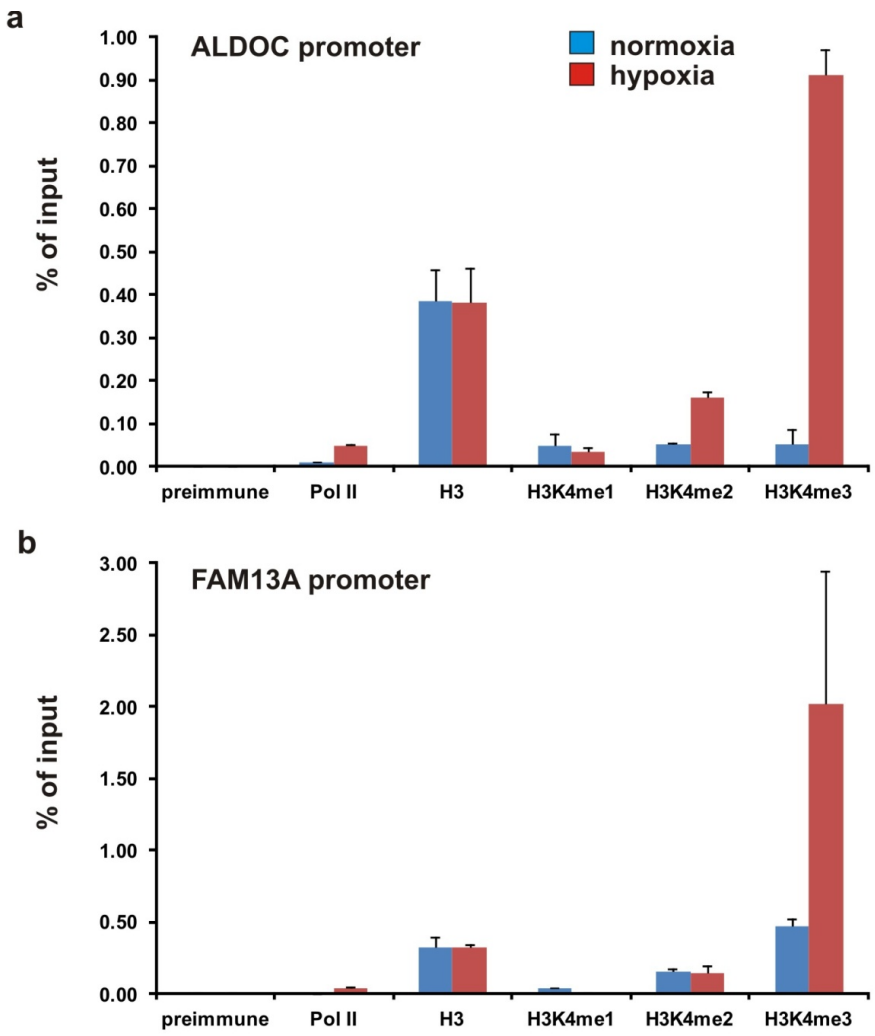


Figure 3.2: ChIP qPCR experiment in MCF-7 cells exposed to normoxia or hypoxia (0.5% O_2 for 16 h) showing levels for indicated histone marks or RNA pol2 at promoters of two HIF dependent regulated genes (a) *ALDO C*, (b) *FAM13A1*. Values were normalised to input levels and expressed at % of input. Bars show mean ± s.d. (n=3). Note that levels for total H3 (H3) are comparable between normoxia and hypoxia indicating equal input levels of the ChIP experiments.

3.2.3 Distribution of histone marks and RNA pol2 under normoxic conditions

To further understand the role of these modifications on a genome-wide basis ChIP-seq experiments were performed using chromatin immunoprecipitated with antibodies to H3K4me1, H3K4me3 and RNA pol2 in MCF-7 cells. In addition, mRNA-seq data were used to compare above mentioned epigenetic modifications with the transcriptome.

Previous reports have shown that RNA pol2, H3K4me3 and H3K4me1 are enriched at gene promoters (Barski, Cuddapah et al. 2007). Furthermore, it was shown that the levels

of these marks correlate well with the abundance of transcripts. Therefore, to provide validation for the genome-wide datasets I examined the distribution of RNA pol2, H3K4me3 and H3K4me1 binding activity under normoxic conditions. The frequency distribution of immunoprecipitated sequences 5' and 3' to transcriptional start sites (TSS) was plotted for the 100 most highly expressed genes, 100 silent genes and 100 intermediate genes determined by the RNA-seq analysis (only gene loci with a unique transcriptional start site were considered) (Figure 3.3). This analysis confirmed comparable patterns of binding to those observed by Barski et al. namely that markers of active gene transcription correlate well with expression levels (Barski, Cuddapah et al. 2007). In particular, for RNA pol2 a distinct peak could be determined at the TSS with lower but definite binding 3' to the TSS within the body of the gene, whereas H3K4me3 and H3K4me1 showed a biphasic peak around the TSS (H3K4me1 being more diffusely distributed) (Barski, Cuddapah et al. 2007).

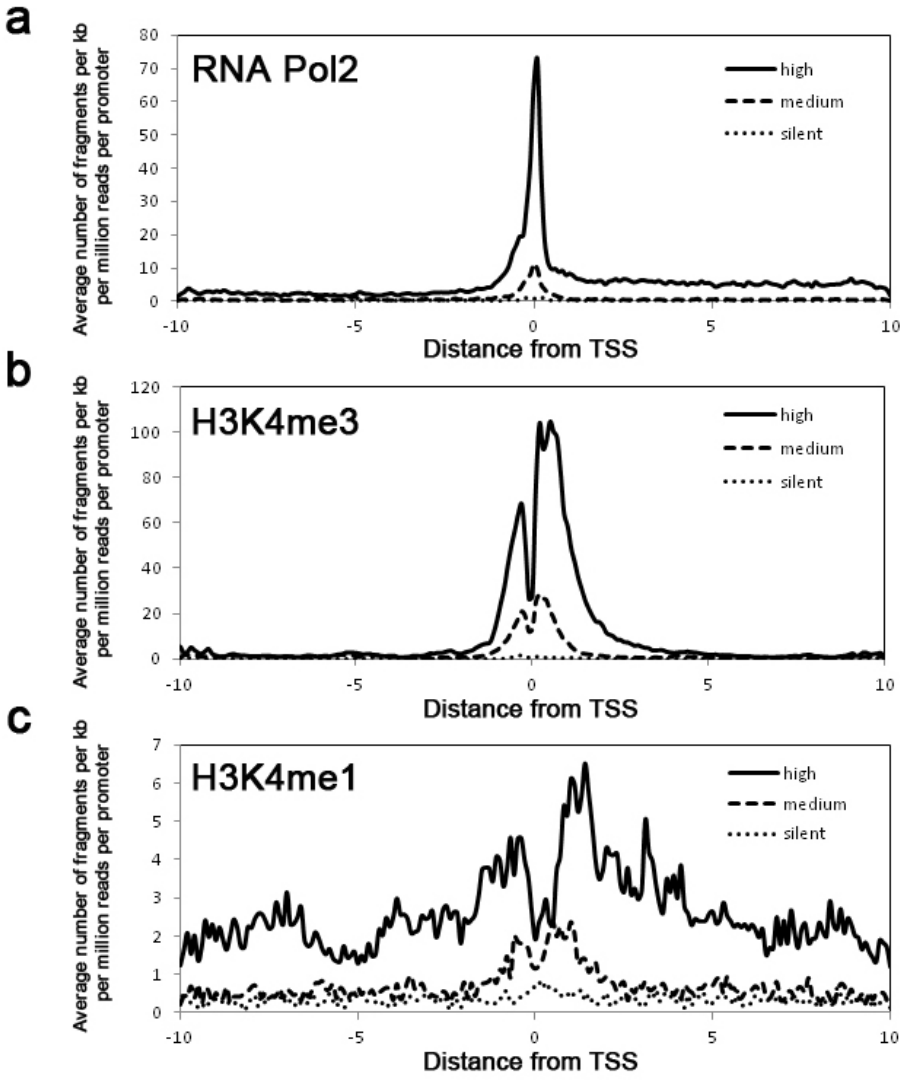


Figure 3.3 Normalised reads for (a) RNA pol2, (b) H3K4me3 and (c) H3K4me1 around transcriptional start sites (TSS) of genes stratified according to their normoxic expression levels in high or medium expressed or silent genes. Reads were not normalised to total H3 explaining the dip for H3K4me3 and H3K4me1 signals at TSS.

In another analysis, the correlation between RNA pol2 abundance either at the promoter or across the body of the gene (without the first 1000 bp) and the transcript abundance was determined across all gene loci with a unique TSS (Figure 3.4). This analysis showed a positive correlation between gene expression and RNA pol2 loading at the promoter or RNA pol2 signals across the gene. Taken together, the analyses from normoxic ChIP-seq data confirm previously described features of epigenetic markers and RNA pol2 in relation

to gene transcription. This represent a good basis for analysing modification of these markers in hypoxia.

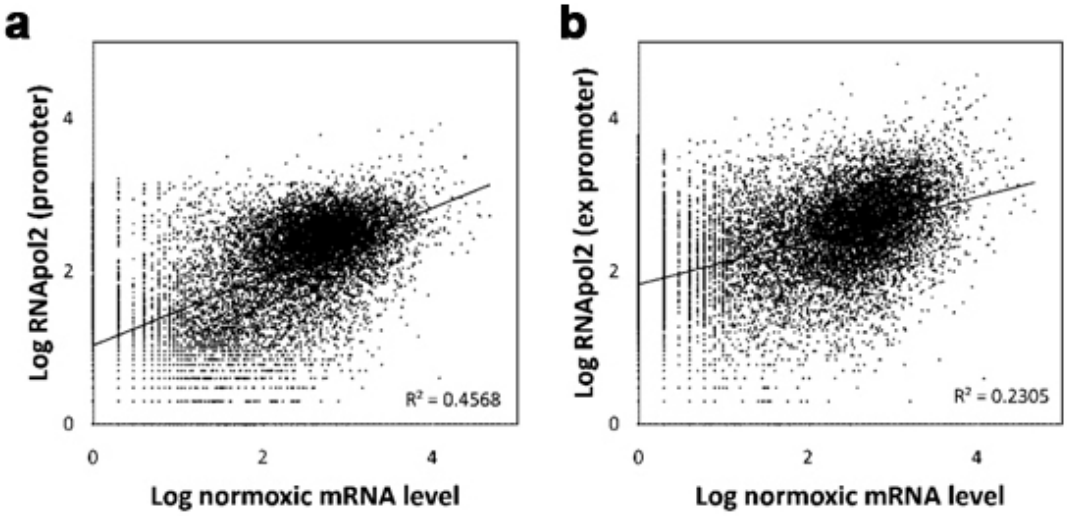


Figure 3.4: Comparison of normoxic mRNA levels with normoxic levels of RNA pol2 at (a) promoters or (b) gene bodies.

3.2.4 ChIP-seq for H3K4me1, H3K4me3 and RNA pol2 under normoxic and hypoxic conditions

3.2.4.1 Changes of RNA pol2 and histone marks at promoters and gene bodies under hypoxic conditions

To directly compare results from histone assays with transcriptional activity RNA pol2 binding was measured by ChIP-seq in parallel. ChIP-seq experiments were performed and reads were visualised using the UCSC Genome browser.

The *EGLN3* locus is shown to illustrate a number of coordinated features seen at individual loci (Figure 3.5). Low levels of H3K4me1 could be detected at the promoter and the HIF-binding enhancer in normoxia, both of which decreased in hypoxia. Conversely, consistent with basal expression of EGLN3, high levels of H3K4me3 could be detected at the promoter in normoxia. However, notably, in hypoxia H3K4me3 levels increased 3' to the promoter, within the body of the gene. Furthermore, in both normoxia and hypoxia a

significant peak could be observed at the HIF-binding site potentially resulting from looping of the enhancer to the promoter even under conditions when HIF is not present. Due to the crosslinking reaction DNA fragments from the interacting enhancer would then be captured by H3K4me3 immunoprecipitation directed at the promoter.

In normoxia, "poised" RNA pol2 was already present at high levels at the *EGLN3* promoter despite low levels of mRNA. In hypoxic conditions the RNA pol2 peak at the promoter remained unchanged but levels of "run through" RNA pol2 increased across the body of the gene and into the 3' region. In addition, at the HIF-binding site a distinct RNA pol2 peak became apparent in hypoxia. Whether this is a result of direct RNA pol2 binding, looping of chromatin to the promoter, or even pausing of RNA pol2 at the intronic enhancer complex is unknown, but is similar to increased levels of RNA pol2 seen at (non-intronic) intergenic HIF-binding enhancers as shown in Chapter2.

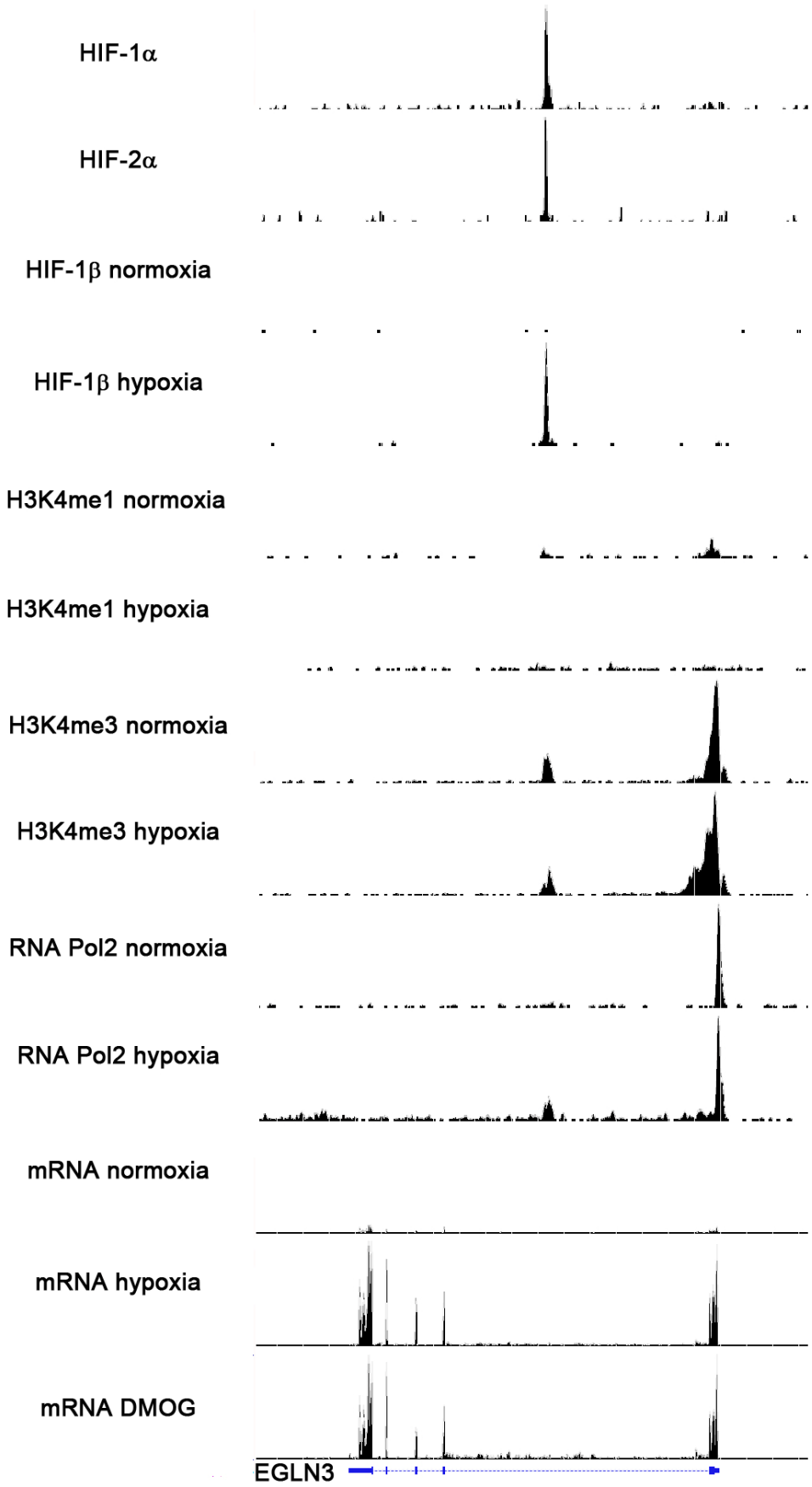


Figure 3.5 ChIP-seq tracks showing read densities for indicated ChIP-seq and mRNA-seq experiments at the *EGLN3* gene locus on chromosome 14.

Many hundreds of genes are regulated by hypoxia and this regulation is largely concordant to that seen with HIF (Elvidge, Glenny et al. 2006). However, whilst similar numbers of genes are both upregulated and downregulated, previous analysis (Chapter 2) has shown that whilst a high proportion of the upregulated genes are direct transcriptional targets of HIF, very few of the downregulated genes bind HIF directly. To test whether epigenetic markers change with hypoxic gene regulation and whether these patterns differ at up versus downregulated genes the distribution of H3K4me1, H3K4me3 and RNA pol2 binding was determined at (a) the 100 most upregulated loci, (b) 100 non-regulated gene loci and (c) the 100 most downregulated loci as determined by the RNA-seq analysis (again only gene loci with a unique TSS were used).

3.2.4.1.1 RNA pol2

Figure 3.6 shows the distribution of RNA pol2 at TSS and the first 10 kb of the gene body. Interestingly, RNA pol2 peaks at the TSS are higher under normoxic and hypoxic conditions at genes upregulated by hypoxia. Furthermore, an increase of RNA pol2 signal can be detected along the gene body of upregulated genes (as seen for *EGLN3* in Figure 3.5), whereas no significant change can be observed at non-regulated genes. Surprisingly, no significant changes in RNA pol2 binding were seen at the downregulated gene loci either - in particular the levels of RNA pol2 "run-through" across the body of the gene did not fall. This finding that upregulation of gene expression in hypoxia is associated with increased "run-through" of RNA pol2 across the gene body, whilst the reciprocal changes are absent from downregulated genes suggests possible differing mechanisms of regulation of transcript abundance.

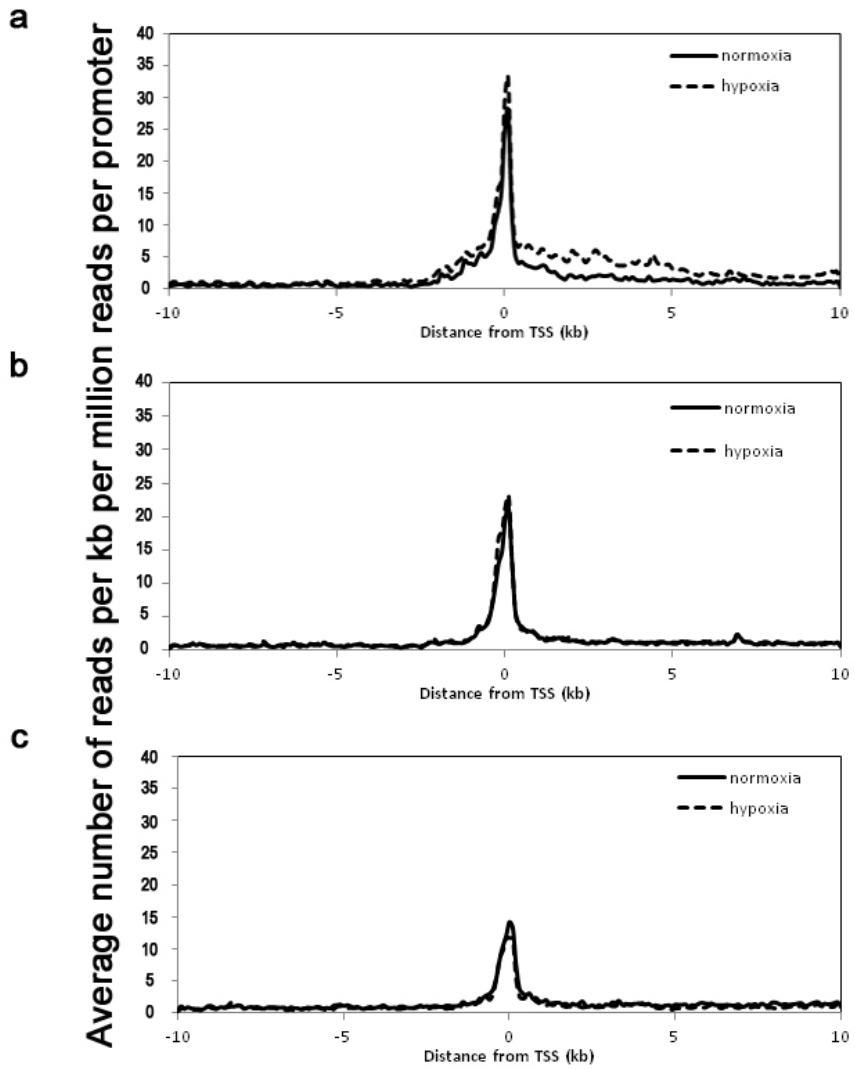


Figure 3.6: Normalised distribution of reads from RNA pol2 ChIP-seq around the transcriptional start sites of genes (a) upregulated (b) not-regulated and (c) downregulated by hypoxia.

3.2.4.1.2 H3K4me3

A similar analysis was performed examining the distribution of H3K4me3 binding close to the transcriptional start sites of the same groups of genes (Figure 3.7). Normoxic patterns of H3K4me3 binding were comparable across all three groups of genes. In hypoxia a broadening of the H3K4me3 signal into the body of the gene was observed, whilst the height of the average peak remained unchanged. No differences between normoxic and

hypoxic patterns of H3K4me3 binding were seen at non-regulated gene loci. Concordant with the normoxic/hypoxic patterns of RNA pol2 binding no changes in H3K4me3 binding were seen in hypoxia at downregulated loci again suggesting differing mechanisms of regulation of transcript abundance.

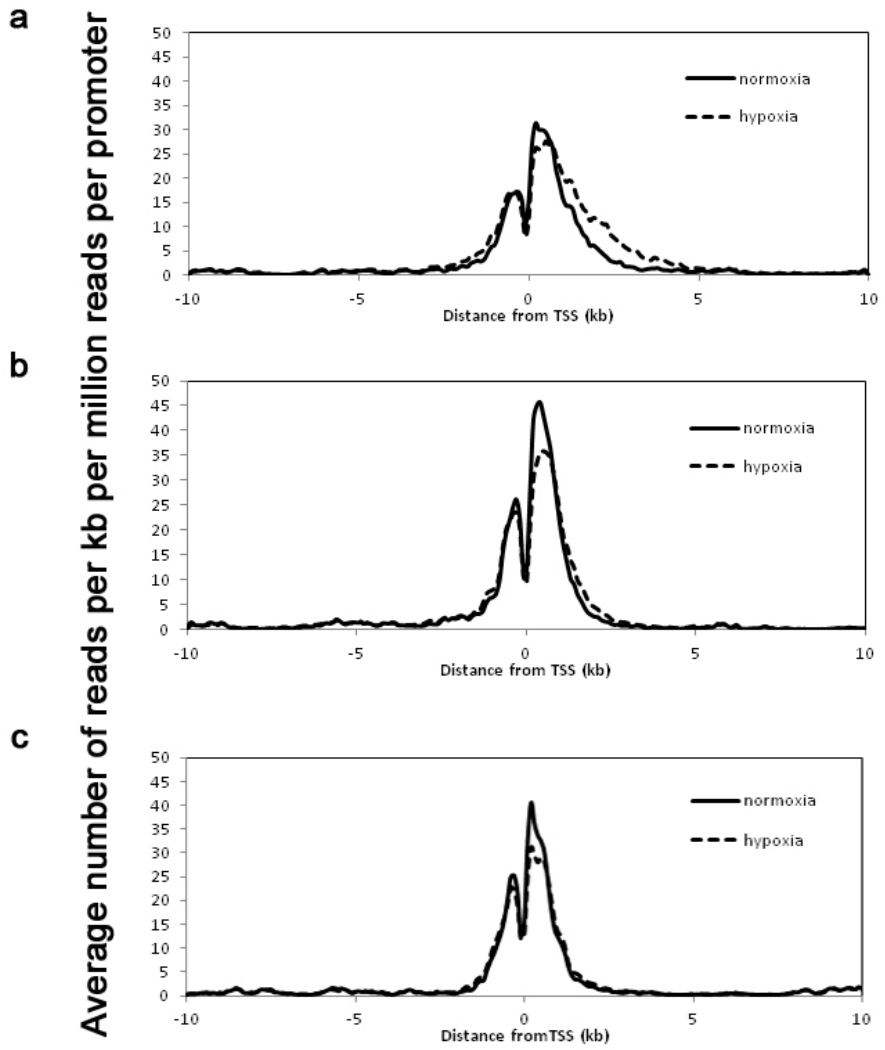


Figure 3.7: Normalised distribution of reads from H3K4me3 ChIP-seq around the transcriptional start sites of genes (a) upregulated (b) not-regulated and (c) downregulated by hypoxia. Reads were not normalised to total H3 explaining the dip for H3K4me3 signals at TSS.

3.2.4.1.3 H3K4me1

In contradistinction to RNA pol2 and H3K4me3 binding, levels of H3K4me1 binding are markedly reduced at all promoters in hypoxia irrespective of regulation of transcript abundance (Figure 3.8). This suggests that changes in H3K4me1 are occurring as a general result of hypoxia and are not regulating or regulated by differences in gene expression.

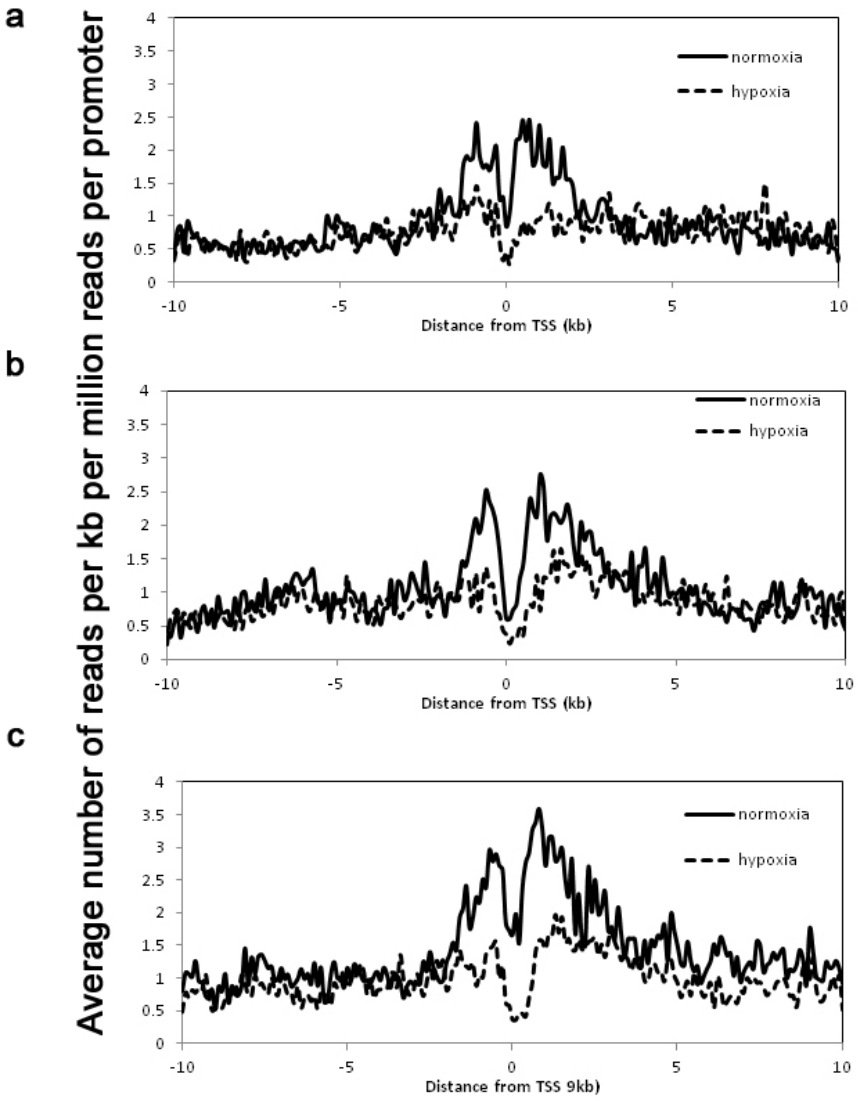


Figure 3.8: Normalised distribution of reads from H3K4me1 ChIP-seq around the transcriptional start sites of genes (a) upregulated (b) not-regulated and (c) downregulated by hypoxia. Reads were not normalised to total H3 explaining the dip for H3K4me1 signals at TSS.

Taken together, the distributions of RNA pol2 and H3K4me1 and H3K4me3 at annotated gene loci and the correlation with gene expression are comparable with published results (Barski, Cuddapah et al. 2007). Major changes under hypoxic conditions appear to occur mainly at genes which are upregulated by hypoxia. RNA pol2 and H3K4me3 signals are increasing into the gene body, which probably reflects increased transcription of these genes (Brookes and Pombo 2009). Importantly, the reciprocal modifications were not observed at downregulated genes, suggesting different mechanisms of regulation of transcript abundance.

3.2.4.2 Changes of RNA pol2 and histone marks at promoters and gene bodies of HIF-binding and non HIF-binding genes upregulated by hypoxia

Since up and downregulated genes differ both in regulation of RNA pol2/H3K4me3 binding and in the frequency of HIF-binding sites, I next determined whether the presence or absence of HIF-binding at upregulated gene loci correlated with different patterns of RNA pol2/H3K4me3 response to hypoxia. Upregulated gene loci were divided into (a) HIF-binding loci, which had a high-stringency HIF-binding site within 100 kb and (b) non-binding loci that had no evidence (no peaks within the top 2000) of HIF-1α, HIF-2α or HIF-1β binding within 1Mb of the promoter. This classification does not exclude the possibility of long-range interactions between HIF-binding sites and promoters, but will largely segregate direct from indirect HIF-targets.

Since a number of factors might confound changes in RNA pol2 binding across the gene locus univariate analysis was performed comparing the RNA pol2 "run through" binding across the body of the gene (number of reads within the transcribed region, excluding the first 1000 bp) with basal RNA level of the transcript, hypoxia RNA level, fold regulation in

RNA and binding status (Table 3.1). Both fold-regulation of mRNA and binding status were correlated with changes in RNA pol2 "run through" across the body of the gene. As a result of this analysis the two groups of HIF-binding and non-binding upregulated genes were matched for fold-regulation of mRNA (Figure 3.9).

Dependent Variable:RNA pol2_exProm

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power a
Intercept	Hypothesis	149.487	1	149.487	1.794	0.396	0.624	1.794	0.092
	Error	89.945	1.08	83.345[b]					
Basal_RNA	Hypothesis	1.732	1	1.732	0.199	0.656	0.001	0.199	0.073
	Error	1671.315	192	8.705[c]					
Hypoxic_RNA	Hypothesis	0.406	1	0.406	0.047	0.829	0	0.047	0.055
	Error	1671.315	192	8.705[c]					
Fold_Reg_RNA	Hypothesis	577.511	1	577.511	66.34	0	0.257	66.344	1
	Error	1671.315	192	8.705[c]					
Binding_Status	Hypothesis	124.972	1	124.972	14.36	0	0.07	14.357	0.965
	Error	1671.315	192	8.705[c]					

a. Computed using alpha = .05

b. .642 MS(Binding_Status) + .358 MS(Error)

c. MS(Error)

Table 3.1: Tests of Between-Subjects Effects

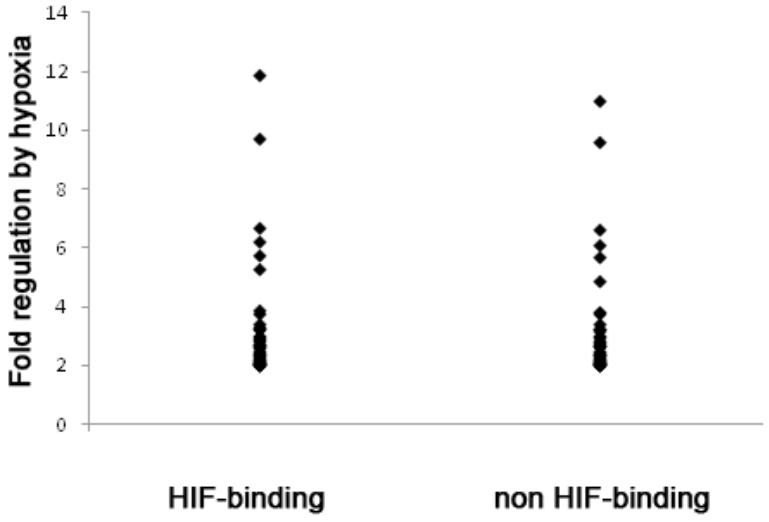


Figure 3.9: Fold regulation of genes selected for analysis of RNA pol2 and H3K4me3 regulation. HIF-binding and non HIF-binding genes have similar distribution of fold regulation.

The distribution of RNA pol2 and H3K4me3 signals around the promoters of these genes were plotted and results are shown in Figure 3.10. Surprisingly, only genes directly regulated by HIF display an increase in RNA pol2 and a broadening of H3K4me3 signals. At non HIF-binding genes no significant difference in signal distribution was observed. This indicates that a subset of genes which is under direct control of HIF was responsible for the increased signals of RNA pol2 and H3K4me3 in hypoxia and that HIF dependent gene regulation is different from other mechanisms leading to gene induction in hypoxia. If, for example, other transcription factors transcriptionally activate non-binding upregulated genes then their mode of action is not via increased recruitment of RNA pol2 and broadening of H3K4me3 binding. It should, however, be noted that ChIP-seq for RNA pol2 only detects the presence of polymerase at the moment of cross linking. It is not possible to calculate transcription/elongation rates from these data and the data is limited to one time point and a certain oxygen level (0.5% in this study). Theoretically, it would be possible that RNA pol2 driven transcription is faster at non HIF-binding genes, therefore leading to similar induction with less total RNA pol2 present at the gene. A possible alternative explanation, which may account for these differing patterns, is that HIF is the major hypoxia-inducible transcriptional activator and that non-binding genes, either upregulated or downregulated, are controlled predominantly through post-transcriptional mechanisms such as RNA-stability. Testing this hypothesis and elucidating possible mechanisms is forming the basis of ongoing work.

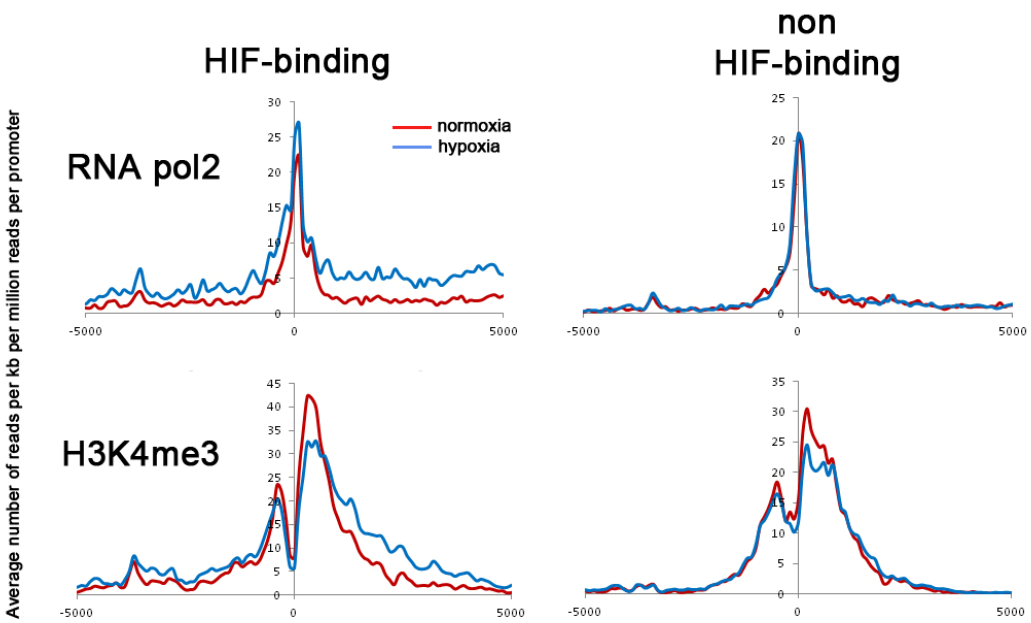


Figure 3.10: Genes with similar amplitude of induction by hypoxia were stratified in two groups: with HIF-binding site or without HIF-binding site in the genomic vicinity. Normalised reads for RNA pol2 and H3K4me3 around the aligned transcriptional start sites of these genes were then plotted for normoxia (red) and hypoxia (blue). Only genes with HIF-binding in the vicinity showed increased levels of RNA pol2 and H3K4me3 towards the coding region of the genes.

3.2.4.3 Changes of RNA pol2 and histone marks under hypoxic conditions at HIF-binding sites distant from promoters

The previous analyses were conducted to test for changes at the promoters of genes regulated by hypoxia generally and by HIF specifically. As outlined in Chapter 2, although HIF-binding is associated with promoters, a great number of functional HIF-binding sites are remote from annotated promoters. In this next section I correlate HIF-binding with changes in RNA pol2 and histone modifications at the HIF-binding sites themselves and distinguish between promoter-proximal and distant sites.

To visualise the effects of hypoxia in the different ChIP-seq experiments, HIF-binding sites were ranked on the vertical axis according to the distance to the next nearest gene promoter (similar as presented in Chapter 2). For each HIF-1-binding site, read densities upstream - left, and downstream - right (direction determined in relation to the orientation of the nearest gene) of the HIF-binding site were then plotted (on a gray scale

- the darker the gray, the higher the read density) for each ChIP-seq experiment (Figures 3.11 and 3.12).

Figure 3.11a shows the positions of CpG islands in the neighbourhood of the HIF-binding sites. Since the sites are arranged according to distance to the closest promoter, the position of the promoters (as outlined by the presence of a CpG island) can be seen in a sigmoid distribution. Those HIF-binding sites distant from annotated promoters are poorly enriched for CpG islands indicating that the majority are genuine long-range enhancers and not unannotated promoters.

Figure 3.11b shows the position of normoxic DNAse1 hypersensitivity sites in the neighbourhood of the HIF-1-binding sites. There is strong enrichment of DNAse hypersensitivity at gene promoters. However, unlike CpG island DNAse hypersensitive sites are enriched at both promoter-proximal and promoter-distant HIF-1-binding sites, indicating that (a) all HIF-1-binding sites irrespective of the distance from the promoter are in regions of open chromatin conformation and (b) these sites are already accessible in normoxia and before HIF-binding - see below.

Consistent with results above, very little HIF-1β binding is observed in normoxia (Figure 3.12a), but is strongly induced by hypoxia (Figure 3.12b) - indicated by the increase in read density at the centre of HIF-binding sites.

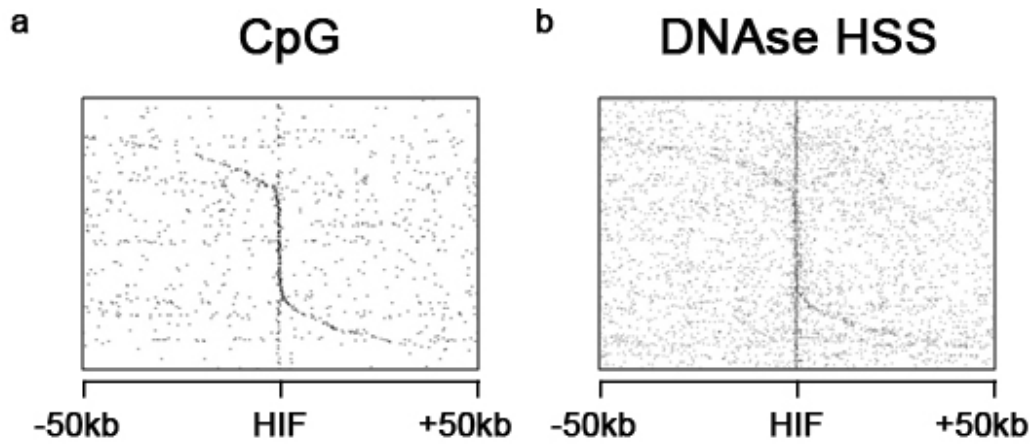


Figure 3.11: HIF-1 binding sites were ranked according to their distance up/downstream of the nearest TSS (vertical axis). The distribution of CpG islands (a) and DNAse HSS sites (b) in relation to (+/-50 kb) the HIF-binding sites is shown.

Concordant with the results above, a peak of RNA pol2 binding is observed at promoters in normoxia (Figure 3.12c). There is also slightly increased signal to the right of these peaks representing "run through" of RNA pol2 across the body of the gene. In hypoxia (Figure 3.12d), this "run through" signal to the right of promoters is increased commensurate with a large proportion of the nearest genes being transcriptional targets of the HIF-binding and with the results above. Importantly, the weak RNA pol2 binding signal observed at promoter-distant HIF-binding sites in normoxia is significantly increased in hypoxia and is consistent with observations at individual loci (*ARRDC3, ERRFI1, ALDOC* and *FAM13A1*). This could represent either direct recruitment of RNA pol2 to these sites, or potentially increased looping of chromatin to allow interaction of the HIF-binding site with RNA pol2 bound at the promoter.

As previously reported H3K4me3 is strongly associated with promoters - with a similar distribution to CpG islands and little enrichment of H3K4me3 is observed at promoter-distant HIF-1-binding sites (Figure 3.12e). In hypoxia, this promoter-associated H3K4 trimethylation extends downstream into the body of the gene (Figure 3.12f) again

consistent with the above findings and with a large proportion of the nearest genes to the HIF-binding site being the transcriptional target of that binding. Although H3K4me3 levels at promoter-distant binding sites are low, they do increase in hypoxia, consistent with looping of chromatin between the enhancer and the promoter.

Under normoxic conditions the signal to noise ratio for H3K4me1 is low (Figure 3.12g) and may represent poor enrichment for this mark by the antibody or more diffuse distribution of H3K4me1 across the genome. However, an association with both, promoter-proximal and promoter-distal HIF-binding sites can be detected. This is in line with reports showing relatively high levels of H3K4me1 at promoters and enhancers (Barski, Cuddapah et al. 2007; Heintzman, Stuart et al. 2007). In hypoxia, H3K4me1 is reduced at the promoters, but is preserved at promoter-distant HIF-binding sites (Figure 3.12h).

Taken together, the results from CpG islands, DNAse1 HSS, RNA pol2 and H3K4me3 confirms the results of the GSEA analysis from chapter 2 that most of the promoter-distant sites are regulatory elements, and have the hallmarks of functional enhancers rather than “cryptic”, previously unannotated promoters.

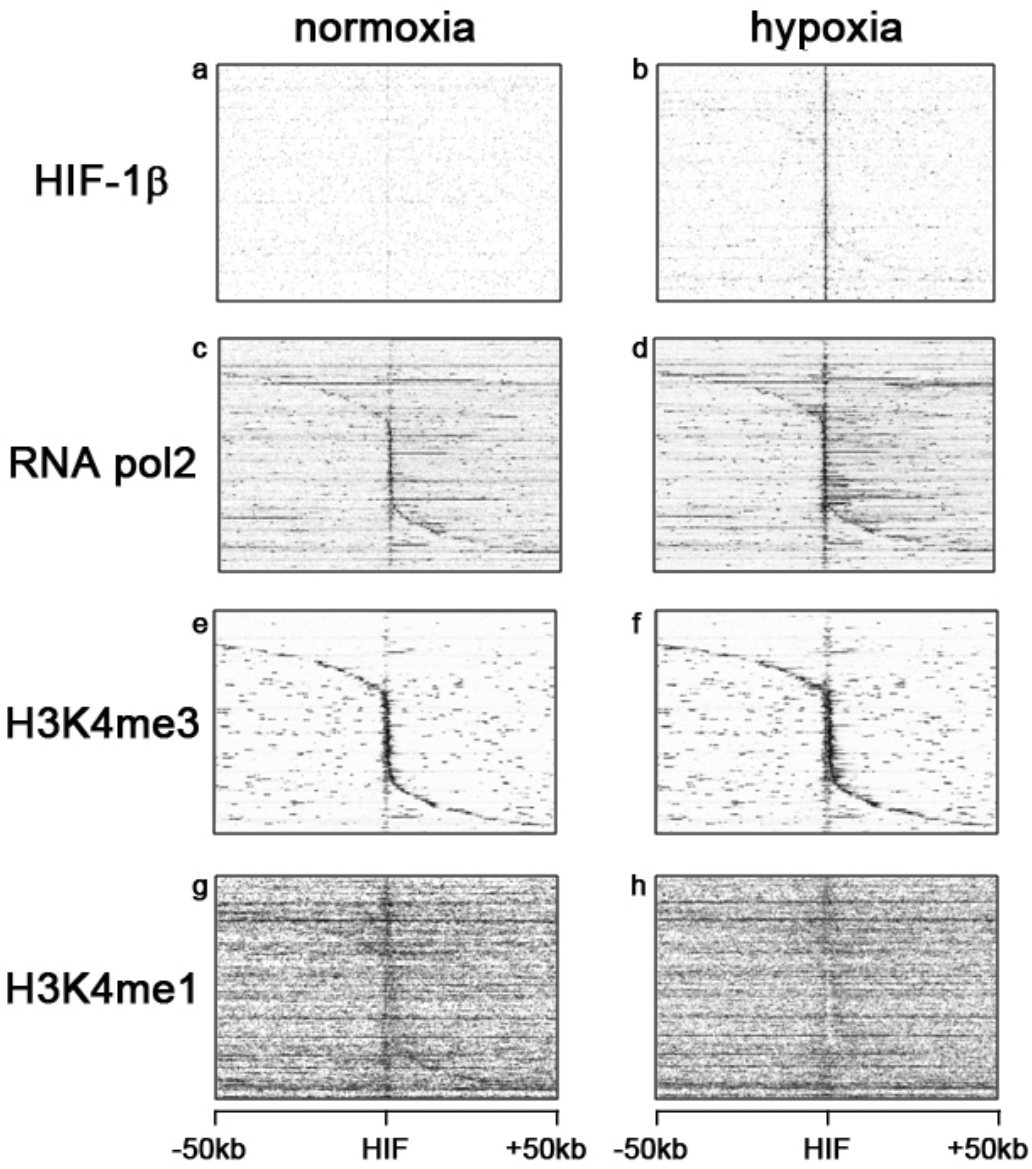


Figure 3.12: HIF-1 binding sites were ranked according to their distance up/downstream of the nearest TSS (vertical axis). Read densities for (a, b) HIF-1β, (c, d) RNA pol2, (e, f) H3K4me3 and (g, h) H3K4me1 in normoxia (left panel) and hypoxia (right panel) illustrate the distribution and enrichment of these marks in relation to HIF-binding sites (+/-50 kb).

To further examine the differences between promoter-proximal and distal HIF-binding in a more quantitative manner, the sites were stratified into two groups (promoter-proximal <1 kb and promoter-distal >10 kb). And the average read density at the HIF-binding sites (±10 kb) for RNA pol2, H3K4me3 and H3K4me1 was plotted (Figure 3.13). Similar to the findings above, signals for RNA pol2 and H3K4me3 were detected mainly at promoter-associated HIF-binding sites and extended to a greater extent across or into the body of

the gene in hypoxia. For promoter-distal sites, increased levels of RNA pol2 were observed in hypoxia, which can reflect either recruitment of RNA pol2 to these sites or more frequent looping to the transcriptional machinery at promoters of their target genes. As expected for enhancers only very low levels of H3K4me3 were observed compared to promoters and these showed a modest induction by hypoxia. Again, this could reflect an increased frequency of looping of the HIF-binding enhancer to the target promoter in hypoxia. Signals for H3K4me1 were as strong at promoters as at enhancers, with only a modest reduction of the signal at enhancers in hypoxia.

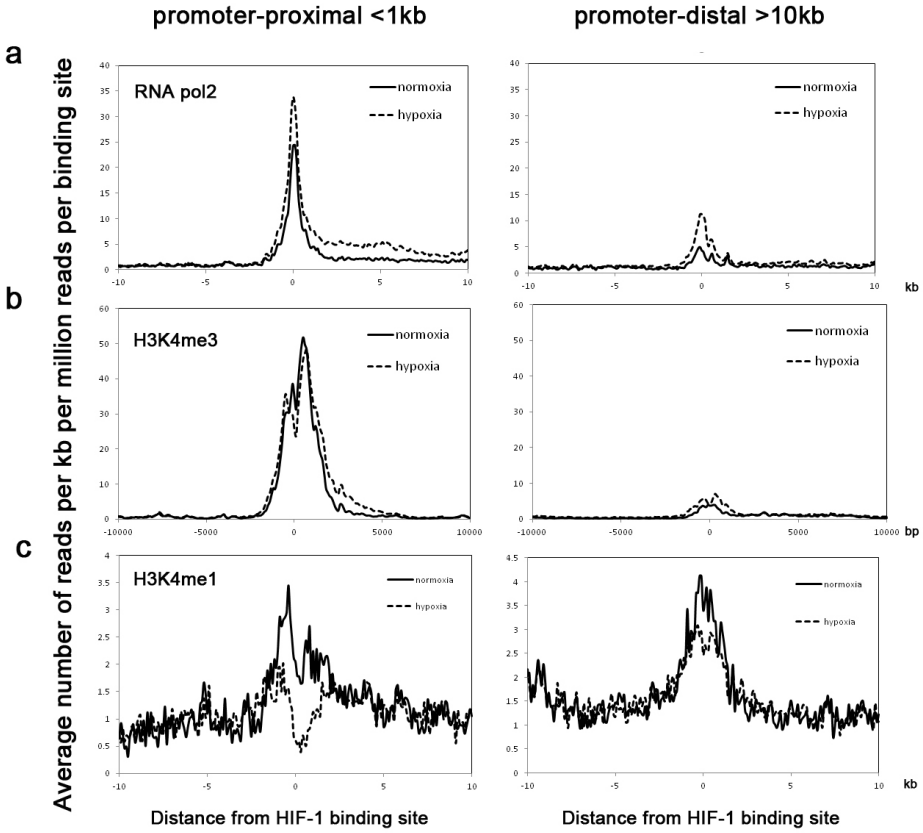


Figure 3.13: HIF-binding sites were stratified in promoter-proximal (left panel) and promoter-distal (right panel) sites and normalised reads for (a) RNA pol2, (b) H3K4me3 and (c) H3K4me1 are shown for normoxia and hypoxia. Reads were not normalised to total H3 explaining the dip for H3K4me3 and H3K4me1 signals at TSS.

In summary, these results confirm findings from Chapter 2, which showed a strong association between promoter-distal HIF-binding and long-range gene regulation,

implying that most of these sites are functional enhancers. Results from ChIP-seq of histone marks describe the presence of enhancer characteristics at the distal binding sites and the increased recruitment of RNA pol2 in hypoxia further supports the functional activity of these sites. Inspection of some of the distal sites in the genome browser, however, revealed that there are a few exceptions from the rule, that promoter-distal HIF-binding is taking place at enhancers (see below).

3.2.5 Detection of non-conventional HIF-binding sites: HIF-1α anti-sense transcript

Genome-wide analysis of epigenetic markers and RNA pol2 binding confirmed the association of promoter-distal HIF-binding with regulatory elements. Together with the functional analysis presented in Chapter 2 the conclusion is that most these sites are enhancers regulating the hypoxic induction of the nearest gene. Although our analyses indicate that HIF rarely acts as a direct transcriptional repressor, there are a number of reports linking HIF-binding to transcript inhibition at individual loci. One such transcript is HIF-1α itself. It has been shown that stabilisation of HIF-1α protein leads to downregulation of HIF-1α transcription (Uchida, Rossignol et al. 2004). Furthermore, inspection of our data sets at the *HIF-1A* locus revealed binding of all HIF isoforms close to the 3' end of the gene (Figure 3.14). Importantly, both ends of the *HIF1A* locus have the hallmarks of a promoter with CpG islands and significant peaks of both RNA pol2 and H3K4me3 binding. In hypoxia there is increased "run through" RNA pol2 across the gene suggesting increased transcription. Concomitantly, there is broadening of the H3K4me3 peak at the 3' end of the gene extending in the 5' direction consistent with increased activity of the antisense 3' promoter at this locus. No such changes were seen at the

conventional 5' promoter. Analysis of mRNA-seq signals indicates downregulation of exonic mRNA levels of approximately 50% consistent with expression array analysis in MCF-7 cells (Uchida, Rossignol et al. 2004; Elvidge, Glenny et al. 2006). Importantly, intronic mRNA-seq signals are strongly increased in both, hypoxia and DMOG treated cells. These results are consistent with the hypothesis that HIF-binding regulates a promoter 3' of the *HIF-1A* gene and induces the expression of an anti-sense (non-spliced) transcript which leads to consequent downregulation of the sense transcript.

Binding of HIF isoforms to this site and HIF dependent regulation of the anti-sense transcript was confirmed in independent samples (Figure 3.15a, b). Furthermore, siRNA mediated knockdown of HIF-1β, which abolishes binding of both, HIF-1α and HIF-2α, rescued the downregulation of spliced exonic HIF-1α transcripts in hypoxia (Figure 3.15c, d). Using primers specific for the anti-sense transcript in qPCR experiments confirmed the induction of the anti-sense transcript, but it should be noted that the mRNA-seq data do not provide proof of the existence of the entire antisense transcript, since the sequencing was not strand specific. However, the strong intronic signals observed in hypoxia and DMOG conditions, the increased RNA pol2 signals and the widening of the H3K4me3 peak, strongly suggest that these signals are amplified from the antisense transcript. However, so far it is not known how the anti-sense transcript is effecting transcription of the sense strand.

In summary, these results indicate that (a) HIF-binding can drive expression of transcripts from unannotated promoters, (b) this may on occasion downregulate the primary transcript and (c) HIF-1α transcription is regulated via HIF-binding in an auto regulatory loop.

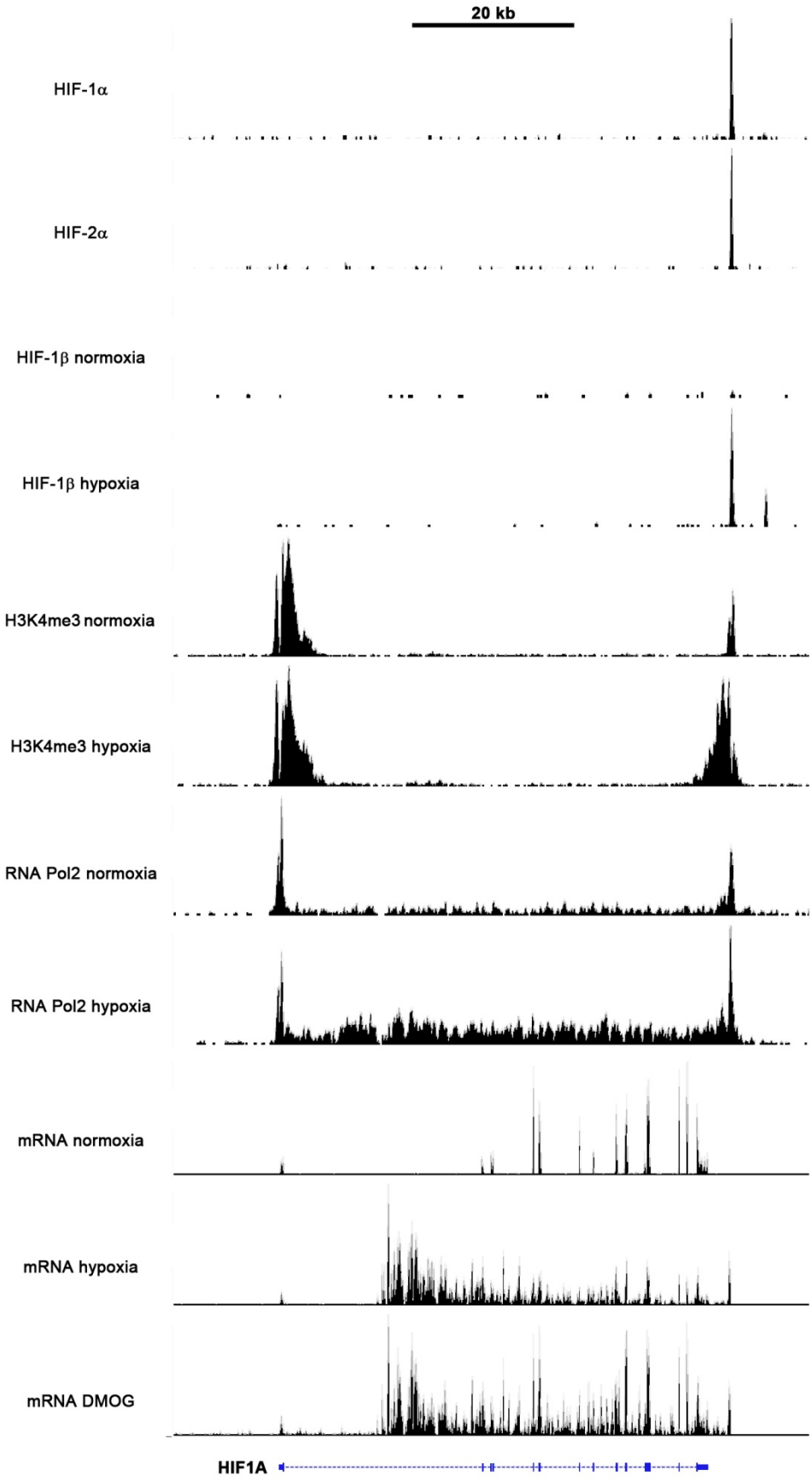


Figure 3.14: ChIP-seq and mRNA-seq tracks at the *HIF1A* locus on chromosome 14. Note the HIF-binding at the 3' end of the gene and the strong induction of RNA pol2 and intronic mRNA signals by hypoxia and DMOG, respectively.

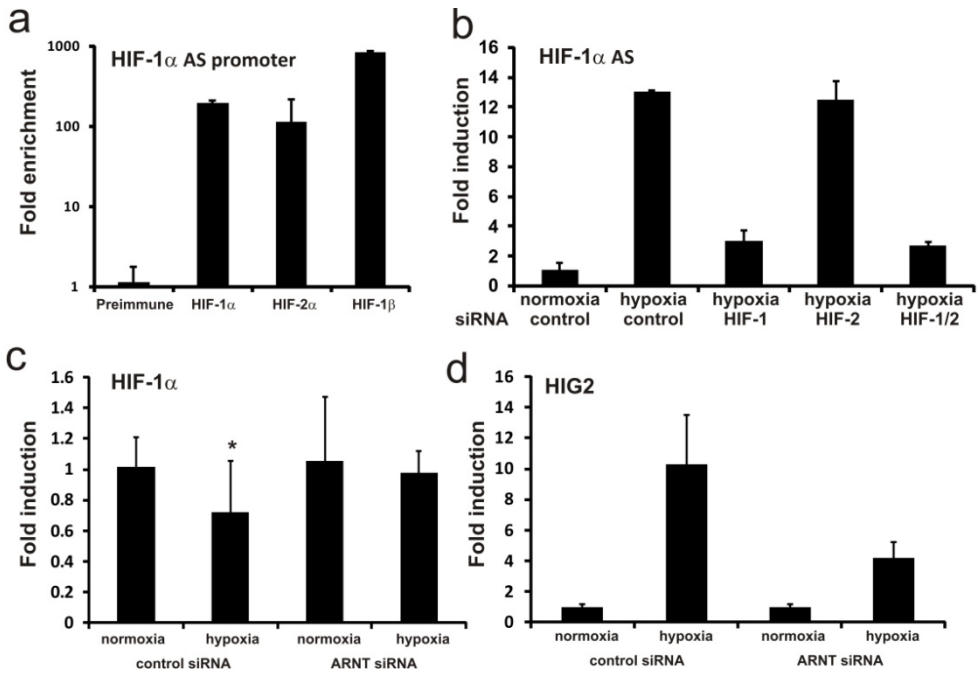


Figure 3.15: (a) ChIP-qPCR confirmation of HIF-binding to the region 3' of HIF-1α in MCF-7 cells (AS: anti sense). (b) Expression-qPCR for the HIF-1α antisense transcript in MCF-7 cells transfected with control siRNA or siRNA targeting the HIF-α isoforms. (c) Expression qPCR for the HIF-1α sense transcript in MCF-7 cells transfected with control of HIF-1β (ARNT) siRNA. (d) Expression of HIG2, an established HIF target in the same samples. Bars show mean ± s.d. (n=3). *) p < 0.05 compared to normoxic control.

3.2.6 Detection of non-conventional HIF-binding sites: a previously unannotated intergenic transcript - "Nici"

Similar to the findings presented above another locus - even more remote from annotated genes (approximately 35 kb upstream of the *SLC2A3* locus) - was identified on chromosome 12, which did not show epigenetic features of an enhancer (Figure 3.16). Inspection of this site revealed the presence of a CpG island together with high levels of H3K4me3 and inducible RNA pol2 binding across a region of approximately 2 kb. mRNA-seq results demonstrated RNA signals in conditions in which HIF is stabilised with only very low levels in normoxia. The structure of the mRNA-seq signals indicates the existence of 2 likely exons. Screening of data bases did not reveal the annotation of any gene in this region of the genome. Therefore, further experiments were conducted to confirm the existence of this novel transcript, named "Nici", and to examine its regulation by HIF.

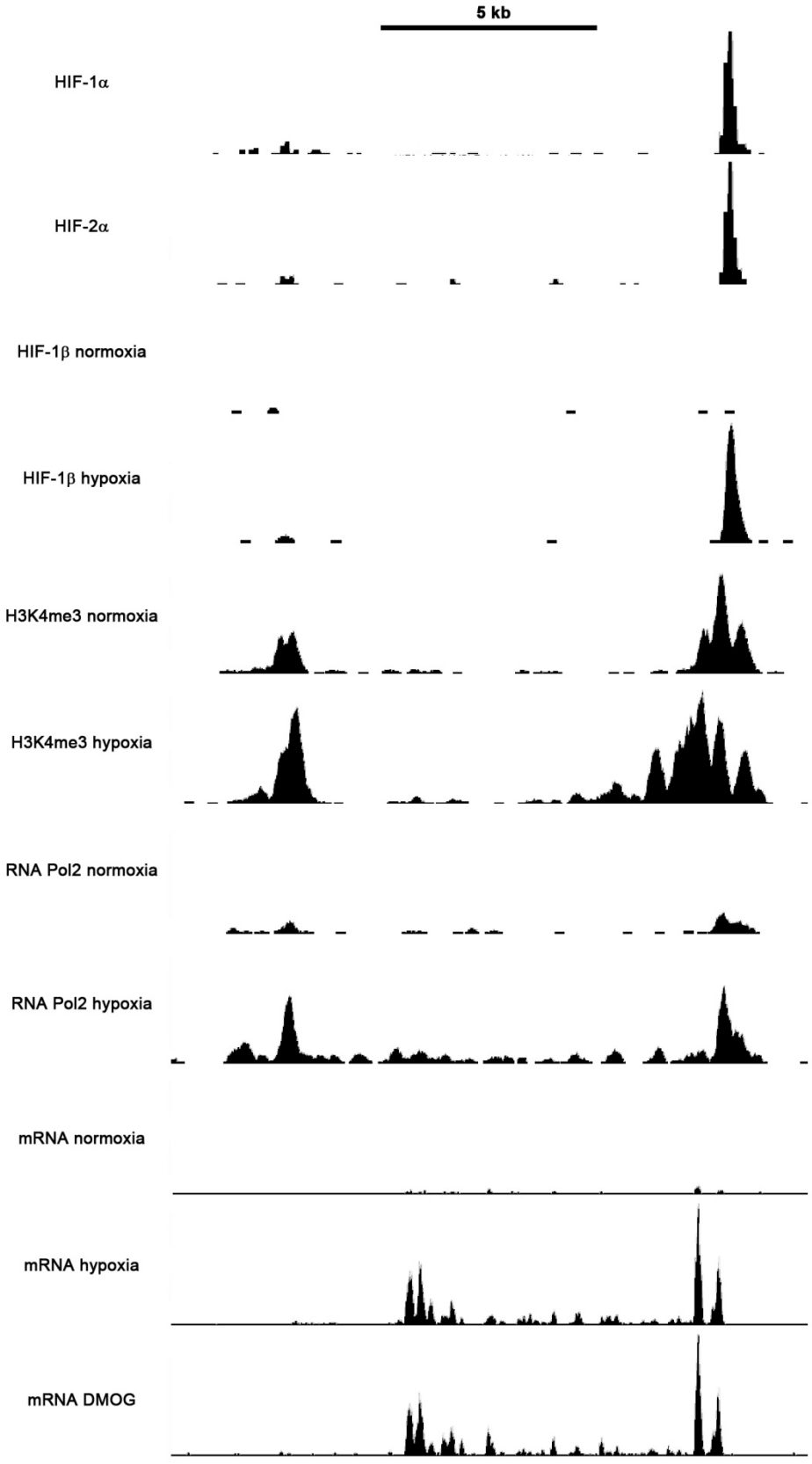


Figure 3.16: ChIP-seq and mRNA-seq tracks at a locus on chromosome 12 remote from annotated genes. Note the HIF-binding to a region with high levels of H3K4me3 and the strong induction of RNA pol2 and mRNA signals by both, hypoxia and DMOG.

First binding of HIF-α isoforms and HIF-1β to this site on chromosome 12 was confirmed using independent samples (Figure 3.17a, b). Similarly, increased RNA pol2 binding to the putative promoter region in hypoxia was confirmed in a further independent set of ChIP samples (Figure 3.17c). Analysis of the HIF-binding site revealed the existence of eight HREs in the putative promoter region. To test the ability of this site to act as a promoter, a sequence spanning the HIF-binding site and containing the HREs was cloned into the pGL3 basic reporter vector. In transfected U2OS cells strong induction of reporter activity by hypoxia was detected in cells containing the "Nici"-promoter construct, which was comparable to the strong induction observed with a *CA9*-promoter construct (Figure 3.17d). To test the existence of mRNA corresponding to the sequences detected in the mRNA-seq experiments, primers were designed covering different regions of the putative transcript. Figure 3.17e shows the induction of well characterised target genes such as *HK2* and *HIG2* by hypoxia or by exposure to DMOG in MCF-7 cells together with the induction of qPCR signals for four primer combinations covering the putative "Nici" transcript. These findings confirm the existence of detectable RNA transcribed from this locus and the induction of this transcript by hypoxia or DMOG. To assess the role of HIF-α isoforms in the regulation of this transcript, siRNA mediated knockdown experiments were performed in MCF-7 cells and cells were exposed to normoxia or hypoxia. Knockdown of HIF-1α almost completely abolished the induction of “Nici” whereas the effect of HIF-2α knockdown was modest (Figure 3.17f). Again this finding is in concordance with the majority of HIF-regulated genes in MCF-7 cells, which are mainly regulated via HIF-1α. Finally, screening of expressed sequence tags (ESTs) and IMAGE clone libraries revealed the existence of a cDNA clone (IMAGE:5549618) from the NIH_MGC_71 clone library which covers this region. The sequence of this clone almost

precisely overlaps mRNA-seq signals in this region (Figure 3.17g, Appendix 8.3, 8.4). It also suggests the existence of a two exon structure and a polyA end of the transcript. The mRNA sequence does not appear to have an open reading frame and therefore it is unlikely that the "*Nici*" locus is a protein coding gene. However, recent reports highlight the existence of a wide range of non-coding transcripts with diverse functions in gene regulation (Esteller 2011). Further experiments such as overexpression of the transcript or knockdown studies should help to identify the functional role of "Nici" in the response to hypoxia.

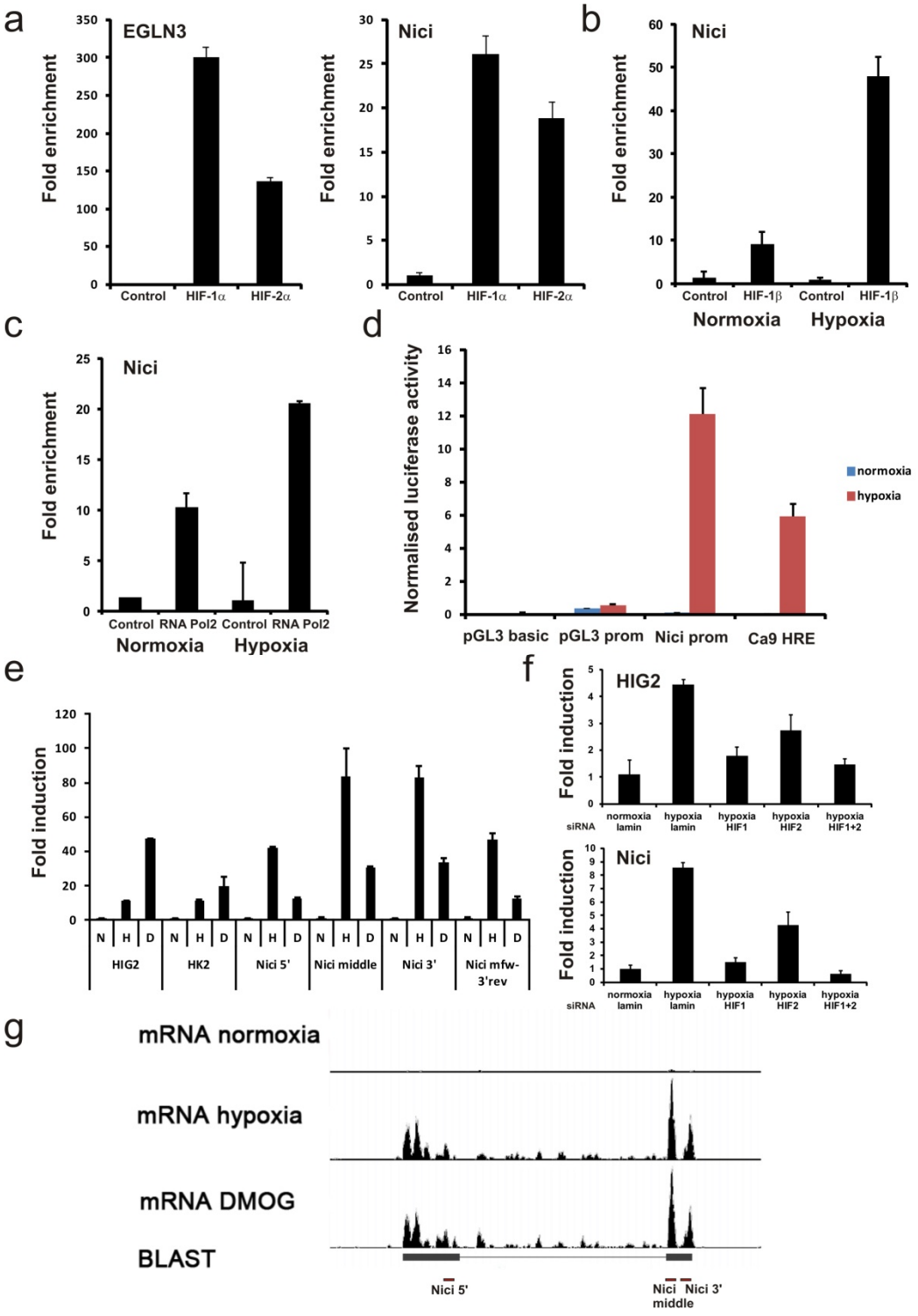


Figure 3.17: (a) ChIP-qPCR confirmation of HIF-α binding to a remote locus on chromosome 12 in MCF-7 cells. Binding to the *EGLN3* intronic HRE was used as positive control. (b) and (c) HIF-1β and RNA pol2 binding is inducible by hypoxia in MCF-7 cells. (d) Reporter assays in U2OS cells exposed to normoxia or hypoxia (1%) and transfected with control plasmids (pGL3 basic, promoter), a sequencing spanning the putative "Nici" promoter and the *CA9* promoter as positive control. (e) Expression-qPCR for HIG2, HK2 and for primer pairs spanning different regions of the putative "Nici" transcript (Primers in (g); MCF-7 cells, N: normoxia, H: hypoxia 1%, D: DMOG 1 mM). (f) Expression qPCR for HIG2 and "Nici" in the same samples described in 3.12b. Bars in all figures show mean ± s.d. (n=3). (g) The putative "Nici" sequence derived from IMAGE clone 5549618 was tested for similarity against the entire human genome using the UCSC browser blast function. The depicted region on chromosome 12 was retrieved suggesting a two exon structure of the transcript consistent with mRNA-seq signals.

3.3 Discussion

In this chapter the genome-wide regulation of RNA pol2 and two selected histone marks (H3K4me1 and H3K4me3) by hypoxia in MCF-7 cells was examined. Signals for RNA pol2 and H3K4me3 in normoxia correlated well with the level of transcription of genes as determined by RNA-seq and in concordance with previously published ChIP-seq studies in CD4+ T-lymphocytes (Barski, Cuddapah et al. 2007). Furthermore, the distribution of reads around the transcriptional start sites of annotated genes was similar to that determined in Barski et al. (Barski, Cuddapah et al. 2007). However, the present study extended the analysis with data generated from cells exposed to hypoxia (0.5% for 16 h). Major findings regarding hypoxic gene regulation were that an increase of RNA pol2 across the gene body and a broadening of the H3K4me3 signal at the promoters towards the coding regions, both indicators of induced transcription, were only observed at gene loci directly under the control of HIF. Upregulated non HIF-binding genes or downregulated genes (which predominantly do not bind HIF either) showed no significant difference in the signal for both, RNA pol2 or H3K4me3.

In Chapter 2 a large proportion of HIF-binding was identified at promoter-distal sites. These sites were characterised by DNAse1 hypersensitivity in normoxia, the lack of overlap with CpG islands (as a marker for promoters) and the functional correlation with hypoxic gene induction at nearby loci. In this part of the study high levels of H3K4me1 and low levels of H3K4me3 (a combination frequently found for non promoter transcription factor binding sites) irrespective of normoxia or hypoxia suggests that most of these remote sites are indeed long-range enhancers of gene expression and that they are already “marked” to be enhancers by the presence of H3K4me1 in normoxia. Consistent with this hypothesis, recruitment of RNA pol2 to these sites in hypoxia confirmed

increased interaction with the basal transcription machinery - another feature that has been described for enhancers (Heintzman, Stuart et al. 2007).

3.3.1 Regulation of histone modifications and RNA pol2 in hypoxia

During the course of this work other studies have been published examining the effect of hypoxia or HIF-stabilisation on selected histone marks in respect to HIF-binding. Tanimoto and colleagues used ChIP-seq to test changes of RNA pol2, H3K4me3, H3ac (acetylated histone 3) and H3K27me3 (a marker for silent chromatin) together with Nucleosome-seq to investigate the epigenomic environment surrounding HIF-binding sites in two different cell lines (Tanimoto, Tsuchihara et al. 2010). Compared to the present study, the authors identified a similar number of HIF-binding sites (531 and 616 HIF-binding sites in DLD-1 and TIG-3 cells, respectively) with a similar distribution with respect to annotated gene promoters. Also consistent with results from this work they found that HIF-binding sites are already in an open chromatin configuration in normoxic conditions and that recruitment of RNA pol2 is a common feature at HIF-binding sites. Xia et al. used a ChIP-on-chip platform limited to promoter regions to examine the status of RNA pol2 and H3K4me3 in normoxia and hypoxia in two different cell lines (HepG2 and U87 cells) (Xia and Kung 2009). In their analyses HIF-1 bound preferentially to genes that were already expressed in normoxia. The clear presence of active markers at the promoters in normoxia (RNA pol2, H3K4me3) and good correlation with levels of transcripts was observed. Again, this is consistent with HIF target sites being “marked” for HIF-binding in normoxia such that the promoters of genes induced by hypoxia have higher levels of normoxic RNA pol2 suggesting a poised state of RNA pol2. These findings would be in line with a model of an “epigenetic template for hypoxic regulation” in cells. This would involve genomic loci (promoter-proximal or promoter-distal) being marked with an open

chromatin configuration and specific histone marks which are prepared for HIF-binding, and gene loci at which promoters are in a poised status marked with RNA pol2 and H3K4me3 and waiting to be activated by the HIF-pathway.

3.3.2 Mechanisms involved in increased methylation of histone 3 lysine 4 in hypoxia

In another study Costa and co-workers exposed A549 cells to nickel, which has been shown to interfere with iron uptake of cells and in addition potentially inhibits HIF-prolyl-hydroxylases directly by replacing iron in the active site leading to a pseudo-hypoxic response (Tchou-Wong, Kiok et al. 2011). Indeed, amongst the top 50 transcripts induced by nickel treatment a remarkable overlap with known HIF targets was observed (e.g. Ca9 and NDRG1 being the top 2). ChIP-seq for H3K4me3 revealed a bulk increase of this mark by nickel treatment and a locus specific increase at induced genes towards the coding region of genes similar to the changes observed in the present study. Costa et al. discussed a potential limitation of ChIP-seq for histone modifications: changes in bulk H3K4me3 levels in immunoblot analysis and locus specific effects on H3K4me3 in ChIP-qPCR appear to be more pronounced than the data derived from ChIP-seq are suggesting, an observation which in the present study was made as well (see Figures 3.1 and 3.2 compared to Figure 3.6) (Tchou-Wong, Kiok et al. 2011). This discrepancy could have technical reasons, since for ChIP-seq equal amounts of DNA from different immunoprecipitations are sequenced after the library preparation step. However, authors noted that after nickel treatment the amount of DNA captured by the H3K4me3 antibody is at least three fold higher than in control samples. This difference would be levelled out in the library preparation step and therefore differences especially at the H3K4me3 peaks would be levelled out as well. This indicates that the true differences between H3K4me3

methylation at specific loci in control or nickel treated conditions (or normoxia and hypoxia in this study) could be even more pronounced than the ChIP-seq analysis suggests. Still, the effect that H3K4me3 is increasing towards the gene body at the 50 most regulated genes by nickel treatment (most of them are HIF-targets) is consistent with findings in this study and suggests that histone 3 trimethylation plays a role in the regulation of gene transcription at HIF-binding sites. In another study Costa and colleagues suggest that inhibition of JARID1A activity by hypoxia (a histone demethylase, which requires molecular oxygen for its activity and which was not induced by hypoxia at the transcriptional level in the cell lines tested) was involved in the increase of global and locus specific H3K4me3 levels and was necessary to drive a specific transcriptional programme (Zhou, Sun et al. 2010). However, in the study by Costa and colleagues a clear link between the effects of JARID1A and HIF-target sites was not presented. Given the fact that several histone demethylases (HDMs, JMJD1B, JARID1A, B, C, D), which are all oxygen-dependent are involved in the demethylation of H3K4 it would seem plausible that the inhibition of a combination of HDMs by hypoxia is involved in the hypoxic response. Indeed, previous studies have shown that hypoxia leads to an increase of bulk H3K4me3 (and H3K9me3 and H3K36me3) levels consistent with findings in this study (Figure 3.1). However, there is little experimental work examining the direct effect of hypoxia on specific HDM activity. Furthermore, the finding that changes in the histone modification H3K4me3, a mark with a very broad distribution across promoters of the genome, are limited to promoters of genes under direct control of HIF argues against a specific HDM or a group of HDMs being inhibited by hypoxia and thereby leading to increased levels of H3K4me3 specifically at HIF-target genes. Indeed it would be more consistent with the specific recruitment of a histone methyltransferase at these sites.

Hypoxia changes transcript levels of many genes that are not under the direct transcriptional control of HIF, but at these gene loci no difference in H3K4me3 levels were observed. In addition, Xia et al reported that increased expression of JARID1B by hypoxia via HIF is compensating for reduced enzyme activity to maintain H3K4 methylation status and suggest a feedback loop similar to the PHD system (Xia, Lemieux et al. 2009). When HIF-1β was knocked down to inhibit HIF-binding an even stronger increase of H3K4me3 levels in hypoxia was detected. The overexpression of JARID1B led to decreased levels of H3K4me3 in hypoxia, indeed pointing to a role of JARID1B in maintenance of H3K4me3 levels in hypoxia. However, again a direct link of JARID1B activity to the expression of HIF-target genes was not examined. For JMJD1A, a HDM which is a HIF target and demethylates H3K9me3 which is linked to gene repression, an interaction with HIF-dependent gene expression was suggested by Krieg and colleagues (Krieg, Rankin et al. 2010). H3K9me3, which was not examined on a genome-wide level by ChIP-seq in the present study, but on the bulk histone level showed increased abundance in hypoxia, consistent with a potential inhibition of JMJD1A by hypoxia. Interestingly, MLL5 a histone 3 lysine 4 methyltransferase which catalyses the reverse reaction compared to HDM was also identified as a HIF target in MCF-7 adding another layer of complexity to this system (Xia, Lemieux et al. 2009; Schodel, Oikonomopoulos et al. 2011).

Taken together, these findings indicate a complex regulatory system of histone methylation in hypoxia involving a potential inhibition of oxygen-dependent HDMs and a feedback loop driven by HIF, which leads to upregulation of specific HDM transcript and protein levels. The effects of these measures appear to be global and, as indicated above, a limitation of the ChIP-seq experiment could be the levelling off of global changes especially at H3K4me3 peaks. Still, the changes towards the coding sequence appear to

be specific for HIF-target genes. The explanation for these increased H3K4me3 levels could be that the increased activity of RNA pol2 at directly HIF-regulated genes leads to recruitment of histone methyltransferases which in turn modify H3 lysine 4 to a higher methylation status (Brookes and Pombo 2009). The escape of RNA pol2 from the promoter to initiate transcription involves phosphorylation of Serin5 (Ser5), which in turn is necessary to recruit the RNA capping machinery modifying nascent RNA. In yeast it has been shown that this modification of RNA pol2 also leads to recruitment of SET1, a histone methyltransferase, which is methylating H3K4me2 to H3K4me3 (Ng, Robert et al. 2003). Increased levels of H3K4me3 are therefore a direct consequence of increased RNA pol2 activity and could be involved in marking the active states of genes facilitating future gene transcription. This hypothesis would be consistent with the observation that changes of H3K4me3 and RNA pol2 are only detectable at gene loci under the control of HIF. The order of events could therefore be that HIF-binding is inducing the release of RNA pol2 from the promoter of target genes. RNA pol2 is Ser5 phosphorylated and recruits, amongst other enzymes, histone methyltransferases which are leading to an increase of H3K4me3. The limitation of increased H3K4me3 levels to the start of the coding sequence could be explained by further events necessary for elongation of transcription such as Serine2 phosphorylation of RNA pol2 which leads to recruitment of other enzymes and changes of other histone marks (H3K36me3 along the body of genes with active elongation).

In conclusion, the regulation of H3K4me3 appears to be complex and could involve direct effects of hypoxia on enzyme activity, a feedback mechanism on HDM expression via HIF and RNA pol2 driven increase of H3K4 methylation status at specific genes. At the moment the most plausible explanation for changes in H3K4me3 levels at the start of the

coding region of HIF-target genes is that increased RNA pol2 activity leads to recruitment of histone methyltransferases to mark the gene as being active. Further studies will be necessary to test this hypothesis in the setting of hypoxia induced gene expression in human cells. This study focused on H3K4 methylation status. Previous reports and this study showed that more modifications are altered in hypoxic conditions involving repressive marks (H3K9me3), markers of active transcription along the gene body (H3K36me3) and that changes in histone acetylation (H3K9ac) are also occurring in hypoxia (Johnson, Denko et al. 2008). In a recent publication a total of 130 post translational modifications on histone residues were identified (67 of which were novel) including eleven known modes of modifications and two novel modifications: tyrosine hydroxylation and lysine crotonylation (Tan, Luo et al. 2011). The latter one was tested for enrichment in ChIP-seq experiments and it could be shown that lysine crotonylation on histones is associated with active transcription. It would be interesting, but of course challenging to extend experiments on histone modifications in hypoxia to this vast number of modifications in order to improve the understanding of mechanisms involved in hypoxic gene regulation.

3.3.3 Regulation of transcript abundance in hypoxia

The fact that HIF is a transcriptional activator and that the changes of RNA pol2 as an indicator of increased transcription are only detected at upregulated genes under direct control of HIF, raises the question how other genes are up or downregulated by hypoxia. As discussed above using ChIP-seq for RNA pol2 only absolute levels of RNA pol2 at one region of the genome at the time of cross linking can be determined, but not the transcription rate per gene. Changes in transcription rate could be responsible for changes in transcript abundance and could escape detection by ChIP-seq. Further

analyses are needed to directly test transcription rate of regulated genes not under control of HIF. Mechanisms independent of transcription are also likely to be involved in regulating mRNA abundance. These mechanisms could involve micro RNAs which as 20-22 bp oligonucleotides are targeting mRNAs for degradation and therefore suppress gene expression. For example, mir-210 which is a well established HIF target has been shown to regulate transcripts related to the cell cycle such as the e2f transcription factor 3 in Hela cells (Giannakakis, Sandaltzopoulos et al. 2008). Mir-210 is highly induced by hypoxia in MCF-7 cells and e2f transcription factor 3 expression is downregulated in MCF-7 cells in a HIF-dependent manner (Elvidge, Glenny et al. 2006). These findings suggest that microRNAs are actively contributing to hypoxic gene regulation in MCF-7 cells. More work is necessary to test the effects of hypoxia on mRNA stability and to identify specific pathways involved in downregulation of genes.

3.3.4 Non-conventional HIF-binding at remote loci: potential functions of HIF-1α antisense and "Nici"

In the present study a strong association of HIF-binding to non-promoter loci with the presence of markers for enhancers of annotated genes was observed. However, some of these remote loci appear to present an exception to this rule. For example, the well documented HIF-1α antisense transcript can be detected in the mRNA-seq data set. The increase is HIF-dependent and the previously reported HIF-binding site is detected in HIF ChIP-seq (Thrash-Bingham and Tartof 1999; Uchida, Rossignol et al. 2004). Uchida and colleagues presented functional evidence that this transcript is responsible for the frequently observed downregulation of HIF-1α mRNA in hypoxia. More recent reports are pointing to an involvement of epigenetic changes evoked by anti-sense transcription at

the *HIF-1A* locus in the regulation of HIF-1α sense transcription (Bertozzi, Iurlaro et al. 2011; Xu, Wang et al. 2012). As another example epigenetic silencing of p15 expression by antisense transcription has been described before (Yu, Gius et al. 2008). In plants, cold induced antisense transcription at a locus required for flowering in Arabidopsis Thaliana has been shown to trigger epigenetic changes to the locus by recruiting polycomb repressors leading to downregulation of the sense transcript (Swiezewski, Liu et al. 2009). It is tempting to apply a similar model to the HIF-system whereby long exposure to hypoxia (or HIF in pVHL-defective renal cancer) leads to increased anti-sense transcription at the HIF-1α locus and over time to silencing of the sense transcript. HIF-1α expression is frequently lost in VHL-defective renal cancer and in most cases genetic alterations are responsible for the decreased expression. In cases in which no such alteration can be detected the possibility of epigenetic silencing could be further evaluated. Furthermore, additional hypoxia induced anti-sense transcripts might exist and a thorough analysis of strand specific mRNA-seq data is needed to identify these gene loci. The fact that HIF-binding is not associated with downregulation of genes makes it unlikely that this is a common mechanism by which HIF regulates gene expression.

Non-coding RNAs were shown to play a major role in regulating gene expression in health and disease (Gupta, Shah et al. 2010). XIST and Hotair are well established examples of non-coding RNAs with important functions in regulation of gene expression. In the present study the second example of remote HIF-binding to a non-enhancer site, the "Nici" locus, associated with a previously unidentified transcript implies that, although not a common feature of the HIF pathway, HIF-binding can induce non-coding RNAs other than micro RNAs. In the p53 pathway a transcript directly induced by p53-binding was identified which has subsequently been shown to be involved in the p53 dependent

downregulation of genes by recruiting repressors to the promoters of these genes (Huarte, Guttman et al. 2010). Interestingly, similar to HIF-binding, p53-binding has been shown to be a transcriptional activator rather than a repressor of gene expression although its pathway reduces expression of similar number of genes. Again it is tempting to speculate that non-coding RNAs are involved in the broad downregulation of genes by HIF and that the "Nici" transcript could be involved in this response. Further experiments are needed to identify the exact role of "Nici" in regard to the hypoxic response. The enrichment for HIF-binding at the promoter is very strong (amongst the top 20 binding sites). The induction of the transcript by hypoxia compared to other direct target genes is very high and levels under normoxia are low (which could explain why this transcript remained elusive for so long and why it is only detectable in unbiased genome-wide mRNA-seq experiments in hypoxia). These facts further support the theory that "Nici" is indeed a functional part of the hypoxic response and that further work is necessary to reveal its biological role in hypoxia.

4 CHAPTER 4: Genome-wide HIF-binding in 786-O renal cancer cells

4.1 Introduction

Kidney cancer accounts for approximately 116,000 deaths per year worldwide and 271,000 new cases are diagnosed annually (Ferlay, Shin et al. 2010). More than 80% of renal cancers are clear cell type tumours (ccRCC) originating from the renal epithelium and most are associated with loss of function of the von Hippel-Lindau (pVHL) tumour suppressor caused by either hereditary - in the context of von Hippel-Lindau syndrome - or acquired gene mutations (Gnarra, Tory et al. 1994) or silencing of the VHL gene (Herman, Latif et al. 1994). pVHL is a ubiquitin ligase and amongst other functions is responsible for initiating the proteasomal degradation of HIF-α subunits under normoxic conditions by recognising hydroxylated proline residues (Maxwell, Wiesener et al. 1999; Cockman, Masson et al. 2000; Bruick and McKnight 2001; Jaakkola, Mole et al. 2001). Loss of pVHL function generates a pseudo-hypoxic transcriptional cascade driven by dimerisation of HIF-α subunits (HIF-1α and HIF-2α) with the aryl-hydrocarbon receptor nuclear translocator (ARNT, HIF-1β) and subsequent binding to hypoxia response elements to activate transcription of hundreds of target genes including *VEGFA*, *CA9* and *GLUT1* (Elvidge, Glenny et al. 2006; Warnecke, Weidemann et al. 2008; Xia, Lemieux et al. 2009). Experimental studies point to a crucial role for HIF-2α over HIF-1α in the development and progression of ccRCCs by enhancing pro-tumourigenic mechanisms (Kondo, Kim et al. 2003; Raval, Lau et al. 2005; Gordan, Lal et al. 2008). A recent report also describes HIF-1α as a tumour suppressor and elegantly provides evidence for the role of genetic deletions of HIF-1α in renal cancer (Shen, Beroukhim et al. 2011). These findings fit well with previous observations in xenograft studies in which the re-expression of HIF-1α has protective effects on tumour growth in 786-O HIF-1α negative cells (Kondo, Kim et al. 2003; Raval, Lau et al. 2005).

Therapeutic options for the treatment of renal cancer so far have been limited, since it is especially resistant to radiation and conventional chemotherapy. Newer therapeutic strategies for treating ccRCCs are rationalised around either inhibiting HIF translation (e.g. mTOR inhibitors) or blocking the action of HIF-activated pathways (e.g. anti-VEGF therapies and tyrosine kinase inhibitors) (Kaelin 2009; Finley, Pantuck et al. 2011; Hutson 2011) although they also have additional actions. However, in the light of the differential effects of the different HIF-α isoforms, therapies that target the HIF system globally might not provide the optimal strategy. Therefore, there is a need to better understand pathways regulated by the different HIF isoforms in renal cancer in order to screen and test for new additional therapeutic options.

The importance of the HIF-2α isoform in RCC biology prompted me to map HIF-2α and HIF-1β binding sites in VHL-defective renal cancer cells that express the single isoform HIF-2α using chromatin immunoprecipitation coupled to high-throughput sequencing in order to identify the pan-genomic pattern of HIF-binding in renal cancer. In addition, expression analyses were performed to determine the functional effects of this binding.

4.2 Results

4.2.1 HIF-binding in different renal cancer cell lines

In preliminary experiments a selection of RCC cell lines with different patterns of VHL and HIF-isoform expression was chosen to test for suitability in ChIP experiments. MCF-7 breast cancer cells from which the previous genome-wide data sets were generated were included for comparison. Caki-1 cells express wild type VHL and therefore retain hypoxic inducibility of the HIF-α isoforms by prolyl hydroxylase inhibition. RCC-4 and 786-O cells are pVHL-defective and therefore stabilise HIF-α in normoxia. In addition, 786-O cells have lost functional HIF-1α due to a truncated transcript and only express functional HIF-2α (Figure 4.1).

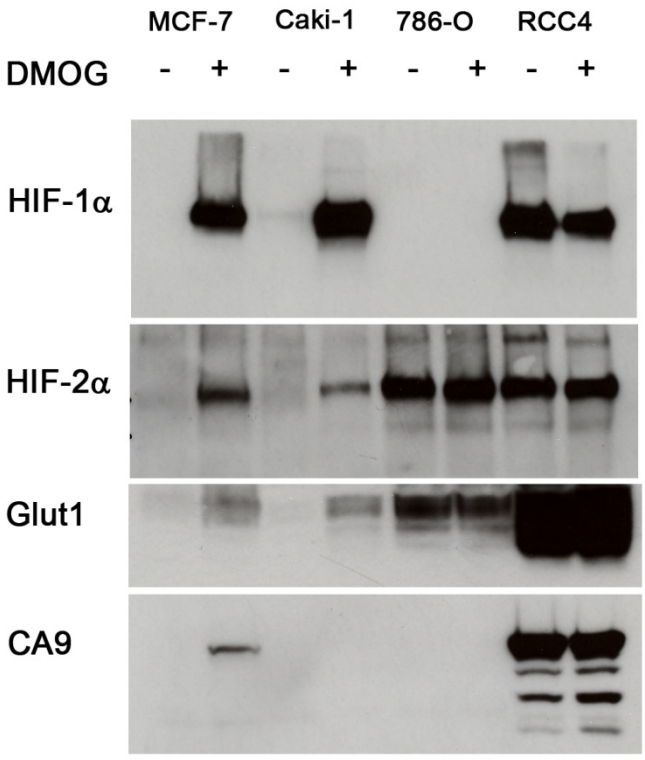


Figure 4.1: Immunoblot analysis for HIF-α isoforms and HIF target genes in different cell lines exposed to 2 mM DMOG or vehicle for 16 h. MCF-7 breast cancer (compare Figure 2.1) and Caki-1 renal cancer cells express functional VHL, whereas in 786-O and RCC4 VHL is functionally lost. Note that HIF-2α is present under control conditions in RCC4 and 786-O cells and that in 786-O cells HIF-1α is not expressed. Note also the different pattern of expression of the target

A selection of HIF-binding sites derived from previous ChIP-on-chip studies (Mole, Blancher et al. 2009) was used initially to screen the renal cancer cell lines for the ability to bind HIF-α isoforms at these loci. HIF-1α and HIF-2α anti-sera were employed in ChIP experiments as mentioned in Chapter 2 and enrichment levels were measured by qPCR. For Caki-1 cells, HIF-α isoforms were stabilised by prolyl hydroxylase inhibition using DMOG, whereas in RCC4 and 786-O cells no treatment was performed. Most of the selected regions showed a degree of enrichment for the HIF-α isoforms. Similar to findings in MCF-7 cells (as described in Chapter 2) the correlation between HIF-1α and HIF-2α binding in Caki-1 and RCC4 cells appears to be very high. Consistent with the lack of HIF-1α in 786-O cells, significant enrichment is seen in this cell line exclusively for the HIF-2α antiserum (Figure 4.2).

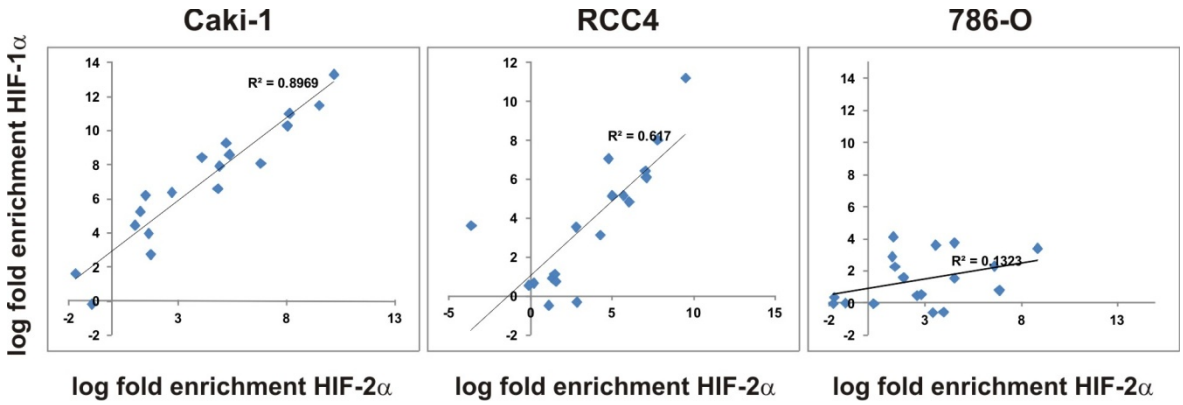


Figure 4.2: ChIP qPCR results for HIF-1α and HIF-2α in Caki-1, RCC4 and 786-O cells. Values were normalised to results from a control region (beta actin) and preimmune serum. Log2 fold values from three PCR replicates from one ChIP experiment and correlation coefficients (r^2) are shown.

4.2.2 HIF-binding in 786-O RCC cells

Since HIF-2 promotes renal cancer progression (and HIF-1 slows growth of RCC xenografts) and since 786-O cells solely express HIF-2α they are a useful tool with which

to determine HIF-binding sites that are relevant for tumour development and progression since a potentially confounding role of the HIF-1α isoform can be excluded. Preliminary comparison of selected HIF-binding sites in 786-O and MCF-7 cells suggests significant overlap of HIF-2 binding events between cell lines but not of HIF-1 binding (Figure 4.3). However, there is considerable scatter suggesting that some sites maybe unique to each cell line. This is in concordance with previously published data from an expression meta-analysis showing that HIF target genes are overlapping between different cell lines (Ortiz-Barahona, Villar et al. 2010) and from gene by gene analyses for single loci (*PGK-1*, *CA9*, *EGLN3*). To further understand whether HIF-binding in renal cancer cells in which HIF is present constantly is different to other cell settings on a genome-wide level and whether this affects other pathways specific for renal cancer, I undertook a genome-wide examination of HIF-binding sites in 786-O cells. As an independent confirmation genome-wide HIF-1β binding in this cell line was examined in parallel.

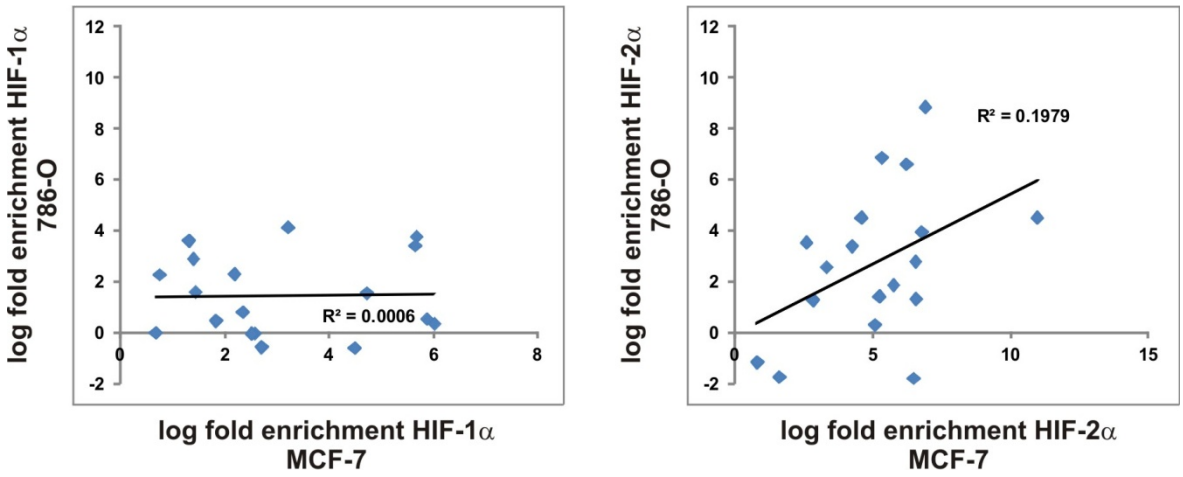


Figure 4.3: ChIP qPCR results as in Figure 4.2 for HIF-1α and HIF-2α in 786-O cells are compared to results from MCF-7 breast cancer cells. As expected there is no correlation between HIF-1α binding in MCF-7 and 786-O cells which are not expressing HIF-1α, but a weak correlation of HIF-2α binding. Values were normalised to results from a control region (beta actin) and preimmune serum. Log2 fold values from three PCR replicates from one ChIP experiment and correlation coefficients are shown.

4.2.3 ChIP-seq for HIF-2α and HIF-1β in 786-O cells

In order to better understand the direct transcriptional targets of HIF-2 in renal cancer, a genome-wide analysis of HIF-2 binding sites in VHL-defective 786-O cells was undertaken by employing chromatin immunoprecipitation using antibodies directed against HIF-2α and its heterodimerisation partner, HIF-1β, coupled to next-generation high-throughput sequencing as outlined above and in the methods section. Data analysis and inspection of the *EGLN3* locus revealed a single, well defined peak for both HIF-2α and HIF-1β at the *EGLN3* intronic HRE (Figure 4.4) coinciding with peaks detected in HIF ChIP-seq experiments in MCF-7 cells.

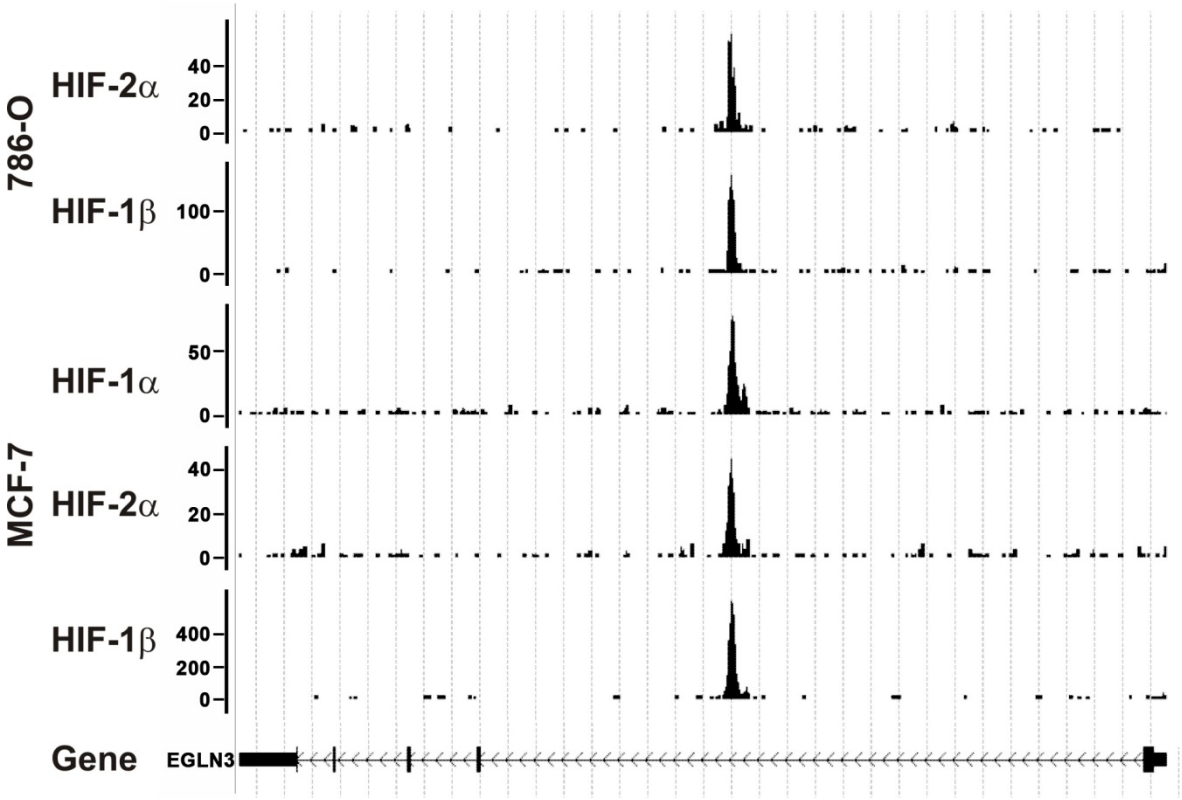


Figure 4.4: Read density tracks from ChIP seq experiments for HIF-2α and HIF1β in 786-O cells and HIF-1α, HIF-2α and HIF-1β in MCF-7 cells across the entire *EGLN3* gene. HIF-binding is detected in all samples at the previously described intronic HRE site. Note that the accumulated number of reads is higher in the MCF-7 HIF-1β experiment due to a higher number of total reads in this experiment.

4.2.4 Overlap of HIF-binding sites between MCF-7 and 786-O cells

To define high stringency HIF-binding sites a similar approach as in Chapter 2 was chosen. A peak threshold of about 28 reads per window was defined using the overlap between 786-O peaks and the previously defined high-stringency MCF-7 peaks (corresponding to approximately 1000 peaks which are still within the steep slope of the overlap before the point of inflexion) (Figure 4.5). This threshold was combined with comparable filters as outlined in Chapter 2 for MCF-7 cells and identified 608 high stringency HIF-2 binding sites (binding both HIF-2α and HIF-1β), at 494 gene loci in 786-O cells (Appendix 8.2.3). A random selection of peaks was confirmed by qPCR using independent ChIP samples from 786-O cells (Figure 4.6).

Defining the degree of overlap between 786-O HIF-2α binding sites and previously identified HIF-1α and HIF-2α MCF-7 binding sites individually revealed remarkably similar concordance for each isoform - 125 (35%) and 108 (36%), respectively (Figure 4.5). This indicates that HIF-1α binding sites in MCF-7 cells are as likely as HIF-2α binding sites to bind the HIF-2α isoform in 786-O cells.

Similarly, the reverse analysis examining the degree of overlap of all MCF-7 peaks with the high-stringency 786-O HIF-2 sites showed a very similar 35% concordance in HIF-binding between the different cell lines with 399 (65%) of the 609 binding loci in 786-O cells being novel sites, with no evidence of HIF-binding in MCF-7 cells (data not shown).

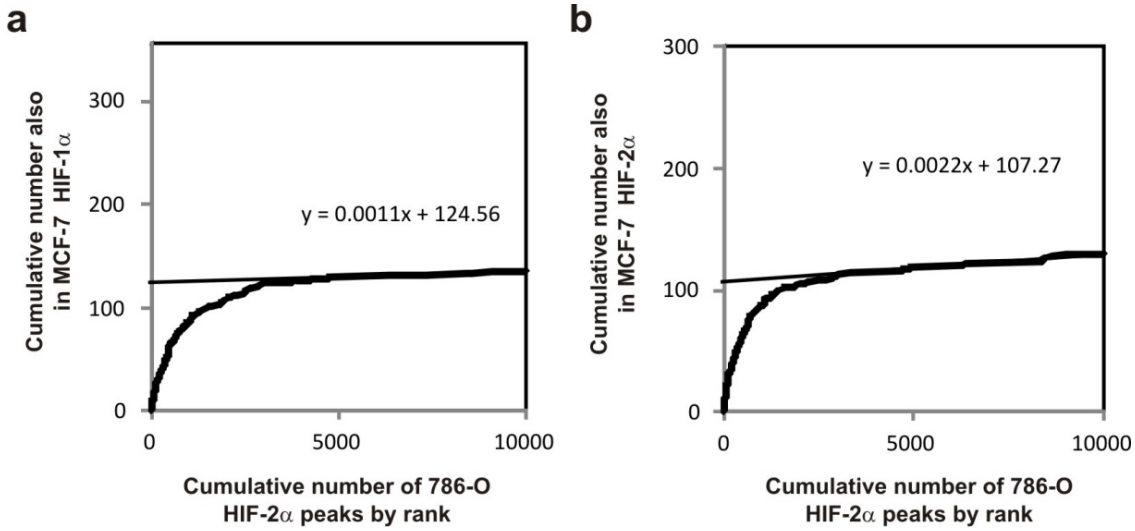


Figure 4.5: Overlap of HIF-2α binding in 786-O with HIF-1α (a) and HIF-2α (b) binding in MCF-7 cells. The x-axis shows HIF-2α peaks in 786-O ranked using CisGenome software according to peak height (number of reads) and the overlap with high stringency HIF-binding loci in MCF-7 cells as identified in Chapter 2 was determined. Beyond approximately 1000 peaks (corresponding to a threshold peak height of 28 reads per window) the degree of overlap fell to background levels. Extrapolating these curves indicates that in 786-O cells HIF-2α binds remarkably similar proportions of the HIF-1 (35%) and HIF-2 (36%) binding sites previously identified in MCF-7 cells.

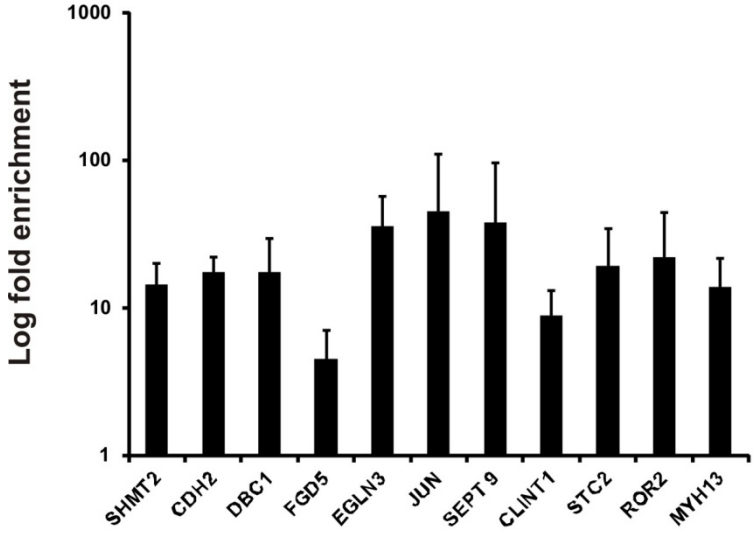


Figure 4.6: Validation of HIF-2α binding in independent ChIP experiments from 786-O cells at randomly selected binding loci derived from the HIF-2α ChIP-seq data set in 786-O cells. Gene names of the nearest annotated gene are given on the x-axis. Mean values from three qPCR replicates of three independent ChIP experiments were normalised to a non enriched control region (beta actin) and to preimmune serum. Bars show mean ± s.d.

4.2.5 Motif analysis in 786-O

Analysis of HIF-binding motifs in 786-O cells (by the same methods used in Chapter 2) did not reveal significant differences in the primary DNA sequence at HIF-binding sites between 786-O and MCF-7 cells indicating that other (possibly epigenetic) mechanisms

such as DNA accessibility account for the differences in HIF-binding (Figure 4.7 and Chapter 2).

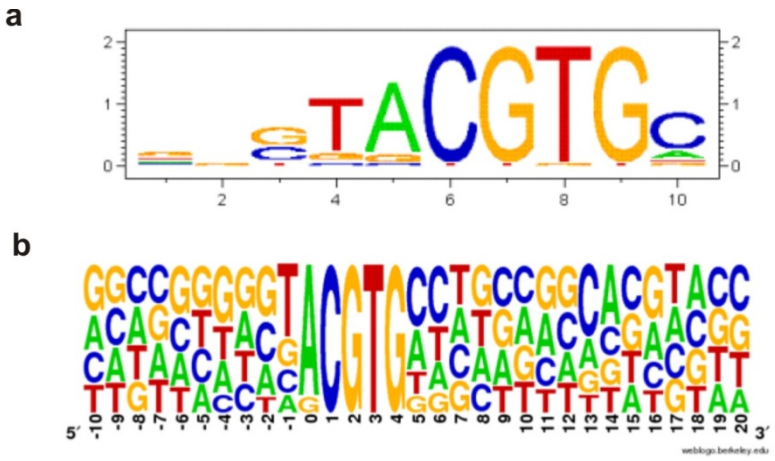


Figure 4.7: (a) HIF-2 binding regions were examined for commonly occurring novel motifs using the Gibbs Motif Sampler module of CisGenome. (b) HIF-2 binding sites containing a single RCGTG motif were analysed for sequence conservation beyond this core motif. Sequences were extended for 100 bp upstream and downstream of the RCGTG motif and aligned. At each position, the observed base frequency was compared with the expected using a Chi-squared test. The relative occurrence of each nucleotide at each position is indicated as a percentage of the total by the letter height.

Analysis of DNAse1 hypersensitivity patterns in MCF-7 cells (ENCODE Digital DNAse1: http://genome.ucsc.edu/cgi-bin/hgTables) indicated that whilst sites common to both cell-types were associated with accessibility in MCF-7 cells, sites unique to 786-O cells were not accessible in MCF-7 cells, despite the nearest promoters generally being accessible (Figure 4.8). Thus local (rather than locus-wide) chromatin conformation is a major determinant of cell-type specific patterns of HIF-binding.

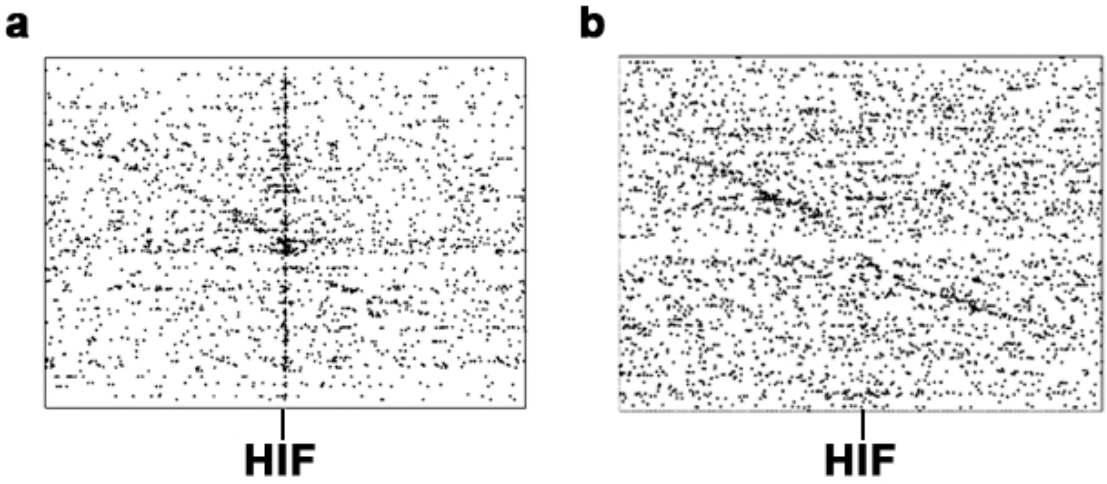


Figure 4.8: The position of DNAse1 hypersensitivity sites determined in MCF-7 cells (ENCODE Digital DNAse1: http://genome.ucsc.edu/cgi-bin/hgTables) was plotted in relation to HIF-binding sites that were (a) common to both MCF-7 and 786-O cells and (b) unique to 786-O cells. HIF-binding sites were ranked according to distance up/downstream of the nearest TSS. MCF-7 DNAse1 hypersensitive sites cluster at the HIF-binding site and at the adjacent promoter (sigmoid curve) at sites common to both cell lines (a), but only at the promoters at 786-O unique sites (b).

4.2.6 Distribution of HIF-binding in 786-O cells compared to MCF-7 cells

One significant finding in MCF-7 cells was that HIF-2 binding occurs more often at promoter-distal sites than does HIF-1 binding. The analysis of the distribution of binding sites around the nearest TSS in 786-O cells, indicated that HIF-2 binding is even less strongly associated with the promoter than either HIF-1 or HIF-2-binding in MCF-7 cells (Figure 4.9a). Subgroup analysis indicated that whilst sites common to the two cell-types showed greater clustering close to the promoter, sites unique to 786-O cells were even more likely to be promoter-distal (Figure 4.9b, c). Thus it is even more unlikely that these novel sites would be identified by classical promoter analysis.

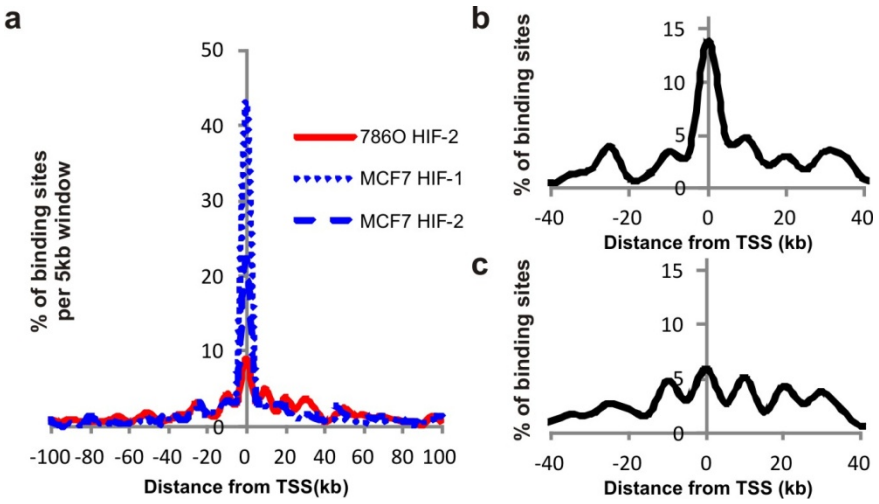


Figure 4.9: (a) Distribution of HIF-binding sides around the transcriptional start site of the next nearest gene. Distribution of binding sites around the TSS that were common to both cell lines (b) or unique to 786-O cells (c). Note that unique sites in 786-O are more likely to be promoter-distant.

4.2.7 Gene regulation by HIF-2α in 786-O cells

In MCF-7 cells transcriptional activity of HIF-2α was found to be very low and only few genes upregulated by hypoxia are affected by HIF-2α siRNA (Elvidge, Glenny et al. 2006). In these cells most of the transcriptional regulation could be ascribed to HIF-1α. To examine the capacity of the HIF system to regulate gene expression in 786-O renal cancer cells, which lack functional HIF-1α subunits, and to identify genes under the direct control of HIF-2 binding I chose two different approaches for genome-wide transcriptome profiling. Firstly, RNA was isolated from 786-O cells transfected with siRNA against HIF-2α or drosophila HIF (dHIF) as control. Secondly, 786-O cells reconstituted with functional VHL (786-O VHL) were exposed to either normoxia or hypoxia (1% O_2) for 16 h before RNA was isolated (Figure 4.10). Both experiments should test the transcriptional activity of HIF-2α in the context of renal cancer.

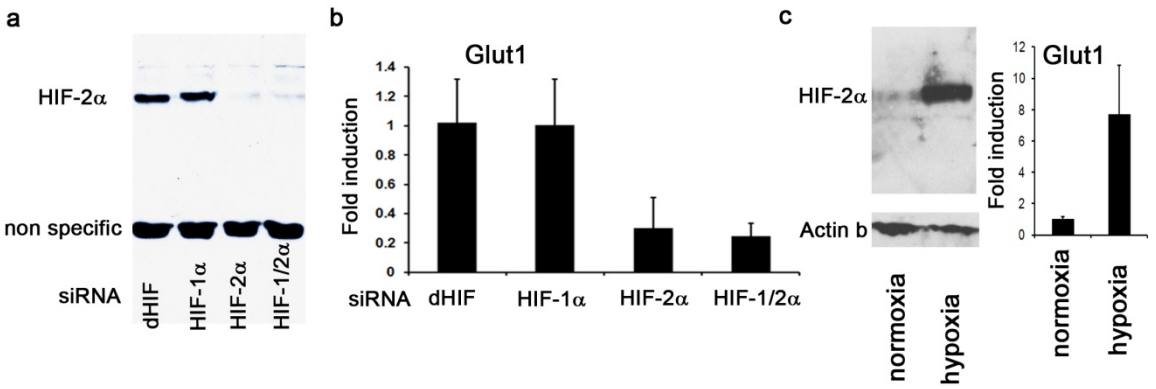


Figure 4.10: (a) Western blot analysis for HIF-2α protein in 786-O cells transfected with siRNA against dHIF (control), HIF-1α, HIF-2α or both HIF-1α and HIF2-α (HIF-1/2α). 786-O cells do not express HIF-1α which therefore serves as another control in this experiment. (b) qPCR for GLUT1 expression in the same samples as in (a). Bars show mean ± s.d. (n=3). (c) 786-O cells reexpressing VHL were exposed to normoxia or hypoxia (1%) for 16 h and HIF-2α protein was detected by immune blot or Glut1 mRNA was measured by qPCR. qPCR values were normalised to expression of an internal control (PPIA) and related to control values using the ΔΔct method. Bars show mean ± s.d. (n=3).

Biological triplicates for each experiment were forwarded to the Microarray Facility at the Wellcome Trust centre and sample preparation, hybridisation to WG-6 bead chips covering approximately 27,500 genes (Illumina, U.K.) and raw data acquisition were carried out by Dr. Dilair Baban at the WTCHG, University of Oxford, U.K. Probe intensities were then quantile normalised and data sets were filtered for expressed genes using a p-value of 0.05 for each probe set as provided by the Illumina software. The relative expression level compared to control conditions was then calculated for each probe-set. Although levels of HIF-2α protein were low and downregulation of the target gene Glut1 detectable in qPCR experiments (Figure 4.10a, b) the maximum fold regulation by HIF-2α siRNA in the array data was modest (maximum of approximately 2.5 fold regulation) in 786-O cells. However, GSEA analysis using high stringency HIF-2α binding sites shows strong enrichment of these at genes upregulated by the presence of HIF-2α (Figure 4.11). The reason for the low fold regulation by HIF-2 knockdown is unclear. Knockdown efficiency was verified by western blotting and appeared to be high. However, in the

context of VHL deficiency even apparently low abundance of HIF-2α could still be functionally highly active and continue to transactivate target genes. To circumvent this potential problem a second approach was chosen in which 786-O cells re-expressing VHL were exposed to normoxia or hypoxia (1%) for 16 h (Figure 4.10c). In the microarray data derived from the 786-O VHL cells, the fold change of expression by hypoxia was more pronounced (maximum of 6 fold regulation) than seen in the previous analysis and again GSEA analysis confirmed clustering of HIF-2α binding sites among genes upregulated by hypoxia (Figure 4.11). These findings validate the functional relevance of the HIF-2 binding sites identified in the ChIP-seq experiment in 786-O cells in respect to gene upregulation and show no association with gene downregulation similar to findings in MCF-7 cells.

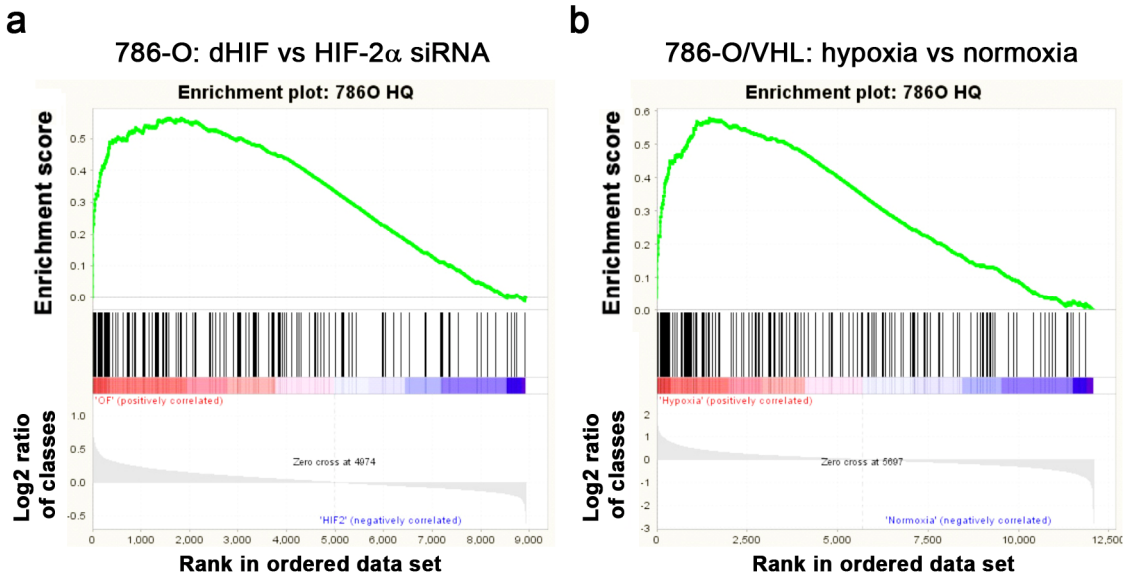


Figure 4.11: Gene Set Enrichment Analysis of HIF-2 binding gene loci in 786-O cells. (a) Expression array data from HIF-2α or control siRNA treated cells (the red area displays genes upregulated and blue area genes downregulated by HIF-2α) or (b) VHL-reconstituted cells exposed to normoxia or hypoxia (the red area displays genes upregulated and blue area genes downregulated by hypoxia) were used, and GSEA analysis was performed. The cumulative enrichment score together with the position of the HIF-binding gene loci amongst the genes ranked by fold regulation are plotted for all high-stringency HIF-2 binding gene loci. The log2 fold regulation by HIF-α is illustrated at the bottom of each panel.

The distribution of HIF-2 binding in 786-O cells is more pronounced at promoter-distal sites than in MCF-7 cells. To test whether these remote regions are transcriptionally

important a comparable analysis to that used for MCF-7 cells was undertaken using the high-stringency (HIF-2α and HIF-1β) binding sites. HIF-binding genes were stratified, according to the distance between the promoter and the nearest HIF-binding site, into 4 groups (less than 1 kb, 1-10 kb, 10-100 kb and more than 100 kb). In all groups, HIF-2 binding sites were enriched among upregulated genes confirming results from MCF-7 cells that HIF-binding is frequently associated with regulation of genes over long genomic distances (Figure 4.12). Given the fact that HIF-2 binding occurs more frequently at promoter-distant sites this finding also raises the possibility that HIF-2α more than HIF-1α is regulating gene expression via binding to enhancers rather than promoters.

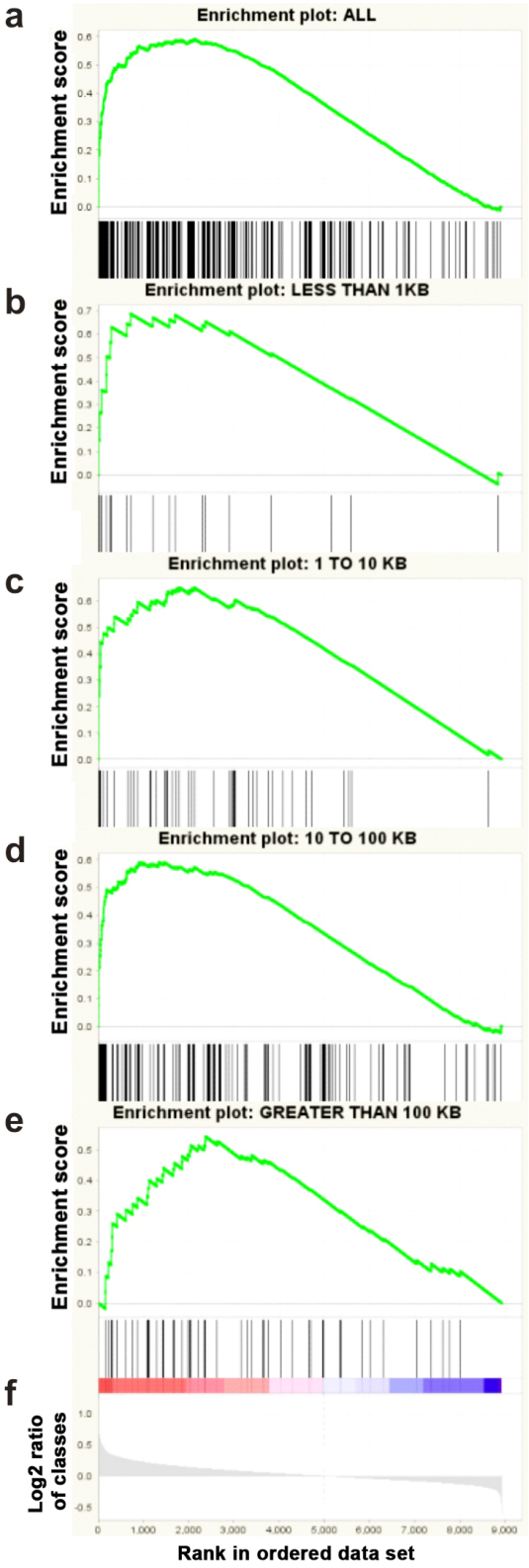


Figure 4.12: Gene Set Enrichment Analysis of HIF-2 binding gene loci stratified by distance between binding site and TSS. Expression array data from HIF-2α siRNA experiments as for Figure 4.11a were used. HIF-2 binding gene loci were stratified according to the distance to the nearest TSS into 4 sub groups and GSEA analysis performed as before. The cumulative enrichment score (ES), together with the position of the HIF-binding gene loci amongst the genes ranked by fold regulation are plotted for (a) all gene loci, (b) gene loci with the closest HIF-binding within 1 kb of the TSS, (c) between 1 kb and 10 kb of the TSS, (d) between 10 kb and 100 kb of the TSS and (e) greater than 100 kb from the TSS. (f) The log2 fold regulation by HIF-2α is illustrated.

4.2.8 Association between HIF-2 binding in 786-O cells and genes deregulated in renal cancer tissue

The previous analysis allows the identification of direct HIF target genes in 786-O cells which are potentially relevant for renal cancer development or progression, but it is limited to the transcriptome of one cell line derived from an individual tumour. Therefore, an analysis was performed testing the extent to which HIF-binding and HIF-dependent gene regulation as defined in 786-O cells can be generalised in the context of renal cancer. Gumz et al. published expression array data sets for ten clear cell renal (ccRCC) tumours, which are deficient for VHL and therefore have high levels of HIF, and for corresponding normal renal tissue in each case (Gumz, Zou et al. 2007). These data sets were used compared with the high stringency HIF-2-binding sites identified in 786-O cells using GSEA analysis (Figure 4.13). Again, HIF-binding sites in 786-O were strongly associated with genes upregulated in ccRCC compared to normal renal tissue. This analysis underscores the relevance of the HIF-binding sites identified in the current study to renal cancer generally and not just 786-O RCC cells since although the use of a single cell line might lead to false negative results, positive calls in both – HIF-binding in 786-O cells and regulation in the setting of renal cancer – strongly indicates genes commonly deregulated in this type of cancer via the HIF-pathway.

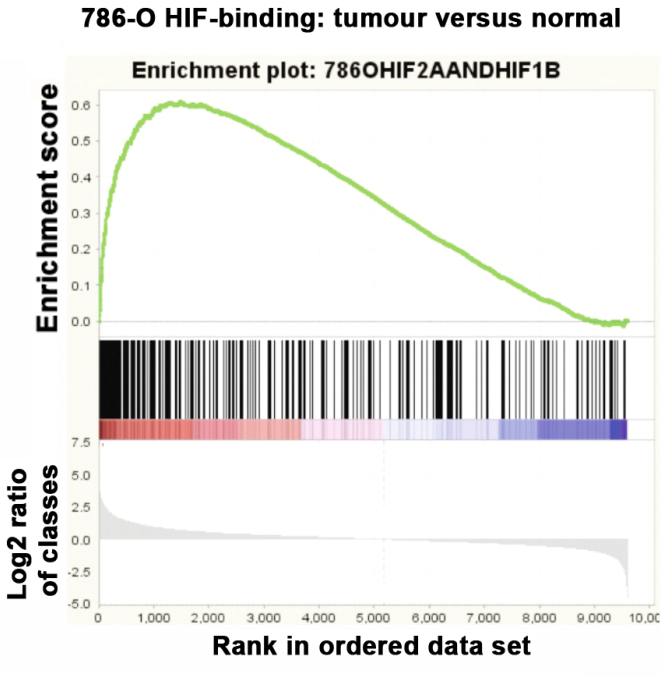


Figure 4.13: Gene Set Enrichment Analysis of HIF-2 binding gene loci stratified. Expression array data from renal clear cell tumours compared to normal renal tissue was used (Gumz, Zou et al. 2007). The cumulative enrichment score (ES, together with the position of the HIF-binding gene loci amongst the genes ranked by fold regulation in renal tumours (red – upregulated, blue – downregulated in renal cancer) are shown.

4.3 Discussion

The major aim of this part of the study is to understand the complex transcriptional response to HIF-binding in renal cancer, to identify candidate binding sites and genes and to explore their relevance for renal tumour development and progression. The genome-wide results presented in this chapter build a basis for subsequent gene-by-gene analyses. Many results confirm the findings from the previous analyses of HIF-binding in MCF-7 cells providing support for their generality. The number of high stringency HIF-binding sites is comparable between the cell lines (approximately 500 in MCF cell vs approximately 600 in 786-O cells). This is also in line with results from a recently published ChIP-seq study on HIF-1 binding in which 531 and 616 binding sites were described in colorectal cancer DLD-1 and embryonic lung fibroblast TIG-3 cells, respectively (Tanimoto, Tsuchihara et al. 2010). However, the comparison of genome-wide analyses in different cell-types has revealed a series of novel findings which provide additional insight into hypoxic gene regulation.

Firstly, the finding that the overlap of HIF-binding between cell lines is more evident at promoter-proximal sites and less prominent at promoter-distal sites points to a crucial role for these more distal sites in fine-shaping the transcriptional activity of genes differentially expressed in different cell lines. In Chapter 2 and in the analysis above evidence was provided that these promoter-distal HIF-binding sites are indeed functionally active irrespective of the distance to the next nearest gene. Gene annotation, epigenetic markers and the lack of CpG islands implies that most of these sites are enhancers of gene expression rather than cryptic promoters. Recent reports underscore these findings and emphasise the modulating role of enhancers in differential gene transcription between cell lines (Heintzman, Hon et al. 2009). With the introduction of

genome-wide analysis of epigenetic markers and transcription factor binding, more information can be gained from these remote regions of the human genome and their involvement in gene regulation.

Secondly, the overlap of HIF-binding sites between different cell types (breast cancer vs renal cancer) was about 35% irrespective of the isoform examined. This is comparable to the degree of overlap reported between different cell lines for other transcription factors (Heintzman, Hon et al. 2009), which further validates the approach chosen for conducting the analysis.

Thirdly, although the fold regulation by HIF-2α siRNA was low, the pronounced transcriptional response of 786-O cells re-expressing VHL to hypoxia indicates that HIF-2α indeed has the functional capacity to induce gene expression on a broader level than previously reported (Elvidge, Glenny et al. 2006; Warnecke, Weidemann et al. 2008). The discrepancy between the fold regulations in the two expression data sets might be caused by technical reasons regarding knockdown efficiency of HIF-2α in 786-O cells or by mechanisms triggered by hypoxia which augment the HIF-2 response in 786-O VHL cells. One such mechanism could be the inhibition of FIH-1 in hypoxia which would lead to increased transcriptional activity of HIF via augmented recruitment of p300. FIH-1 is expressed in the human kidney and is present in RCC cells (Schodel, Bohr et al. 2010; Khan, Bhattacharyya et al. 2011). However, this theory would suggest that transcriptional activity of HIF-2 in non VHL-reexpressing cells is low anyway and that a hypoxic environment is needed for HIF to be active even if VHL is not present. The fact that rexpression of VHL does substantially change the expression of HIF-targets in normoxia speaks against this theory. Furthermore, the ability of FIH-1 to hydroxylate HIF-2α has been shown to be much lower than for HIF-1α (Bracken, Fedele et al. 2006) and a recent

report found no effect of FIH-1 knockdown on HIF-2α induced gene expression in 786-O cells (Khan, Bhattacharyya et al. 2011). However, in other RCC cell types in which HIF-1α is still expressed a tumour promoting function of FIH-1 has been described via inhibition of HIF-1α mediated cell apoptosis (Khan, Bhattacharyya et al. 2011).

The more pronounced HIF-2α effects can be related to the context of HIF-1α deficiency in 786-O cells in which HIF-2α is the only isoform present and therefore could take over some transcriptional activity from HIF-1α or act more efficiently in the absence of HIF-1α. Interestingly, although most of the binding sites are shared by HIF-1 and HIF-2, the finding from Chapter 2 that HIF-2 binding occurs at more distal sites from promoters seems to be confirmed in 786-O cells, in which high stringency binding sites for HIF-2 are even more remote to annotated promoters compared to HIF-1 or HIF-2 binding in MCF-7 cells. The reason for this finding remains unclear. Motif analysis did not reveal differences between HIF-1 and HIF-2 binding sites, indicating that other mechanisms, possibly epigenetic mechanisms are directing the HIF-α subunits to individual binding sites.

Fourthly, another finding is the strong enrichment of HIF-binding sites amongst upregulated genes in renal cancer tissue, which to some extent is predicted since stabilised HIF-α isoforms are a hallmark of pVHL-defective renal clear cell cancers. Although a few HIF-binding sites already described in renal cancer were not detected in 786-O cells (e.g. *CA9*) HIF-binding in 786-O cells in general reflects the HIF driven response in renal tumours very well. Therefore, genome-wide analyses of the HIF-response in these cells should provide a good basis to identify candidate genes with a potential driver role in RCC.

5 CHAPTER 5: Identification of a HIF-binding enhancer of cyclin D1 expression specific for VHL-defective renal cancer

5.1 Introduction

During the course of this work a genome-wide association study in patients with renal cancer discovered two SNPs (rs11894252 and rs7579899) located in a regulatory region of the first intron of the *EPAS1* gene (coding for HIF-2α) on 2p21 significantly associated with higher risk for RCC. This provides the first genetic evidence in humans for the importance of HIF-2α in the development of renal cancer (Purdue, Johansson et al. 2011). In the same study an additional SNP (rs7105934) on 11q13.3 reached genome-wide significance - the minor allele with an allelic frequency of 7% being associated with protection from renal cancer. The association of this locus with renal cancer has now been replicated in two further studies, in a population of European descent and in a Chinese population (Cao, Qin et al. 2012; Wu, Scelo et al. 2012). rs7105934 is located in an intergenic region flanked centromeric by *MYEOV* (myeloma overexpressed (in a subset of t(11;14) positive multiple myelomas) gene) and telomeric by the gene coding for cyclin D1 which is commonly deregulated in multiple types of cancer (Musgrove, Caldon et al. 2011). Interestingly, cyclin D1 has been found to be HIF-dependently regulated only in the context of VHL-defective renal cancer (Bindra, Vasselli et al. 2002; Zatyka, da Silva et al. 2002; Baba, Hirai et al. 2003; Wykoff, Sotiriou et al. 2004; Raval, Lau et al. 2005). Importantly, all disease-associated loci identified in these GWAS analyses lie in non-coding regions that are not included in exome sequencing approaches and therefore suggest a mechanism mediated through regulation of gene transcription, rather than through altered gene function.

In the previous chapter, comparison of genome-wide HIF-binding between different cell lines showed that differences in HIF-binding are more pronounced at promoter-distal sites indicating that such enhancers are involved in shaping the differential responses

evoked by HIF in different cells. This is in concordance with findings in a recent study (Heintzman, Hon et al. 2009) that show that it is differences in enhancer activity rather than promoter activity that largely governs cell-specific gene regulation. Importantly, comparing binding sites in MCF-7 and 786-O revealed a 786-O specific site that matched this model approximately 220 kb upstream of the *CCND1* gene on chromosome 11 and approximately 5 kb upstream of the SNP rs7105943, which is associated with protection from renal cancer. This finding and the restriction of HIF-dependent cyclin D1 regulation to renal cancer cells suggests a potential direct link between HIF-2α binding and cyclin D1 regulation. In the next chapter I will further characterise the HIF-binding site on 11q13.3, and examine its role in the regulation of cyclin D1 expression. Furthermore, this work will present data on how HIF is a major factor involved in defining the renal cancer susceptibility locus on 11q13.3.

5.2 Results

5.2.1 HIF-binding to an intergenic remote region on 11q13.3

As outlined in Chapter 4, differences of HIF-binding between cell lines appear to be more pronounced at promoter-distal binding sites than at promoters. For example, HIF-binding is observed at a remote intergenic region on 11.q13.3 in 786-O renal cancer cells but not in MCF-7 breast cancer cells. Figure 5.1 shows HIF ChIP-seq tracks in this region and illustrates high-stringency HIF-2 binding (HIF-2α and HIF-1β binding both significant on a genome-wide level) in 786-O cells at a major peak (11q13.3 H, ranked 12th by peak height in the 786-O data set) and two minor peaks (11q13.3 h1 and h2, not significant on a genome-wide level), but no signals at any of the three sites in the MCF-7 tracks. This locus on 11q13.3 is located in a gene poor region and is flanked by the gene *MYEOV* approximately 150 kb centromeric and by the gene encoding cyclin D1 (*CCND1*) approximately 220 kb telomeric. Previous reports have shown that regulation of cyclin D1 by the HIF pathway occurs in the context of renal cancer cells but not in other cell types tested (Baba, Hirai et al. 2003; Raval, Lau et al. 2005). Interestingly, knockdown and overexpression studies point to HIF-2α as the main driver of cyclin D1 expression in renal cancer (Raval, Lau et al. 2005). It is unclear whether HIF-2α is directly regulating cyclin D1 expression via a so far cryptic binding site or indirectly via secondary mechanisms. Given the established role of cyclin D1 in tumour biology (Musgrove, Caldon et al. 2011) and the recent genetic evidence for the 11q13.3 region in the development of renal cancer I decided to pursue further studies on this remote HIF-binding site.

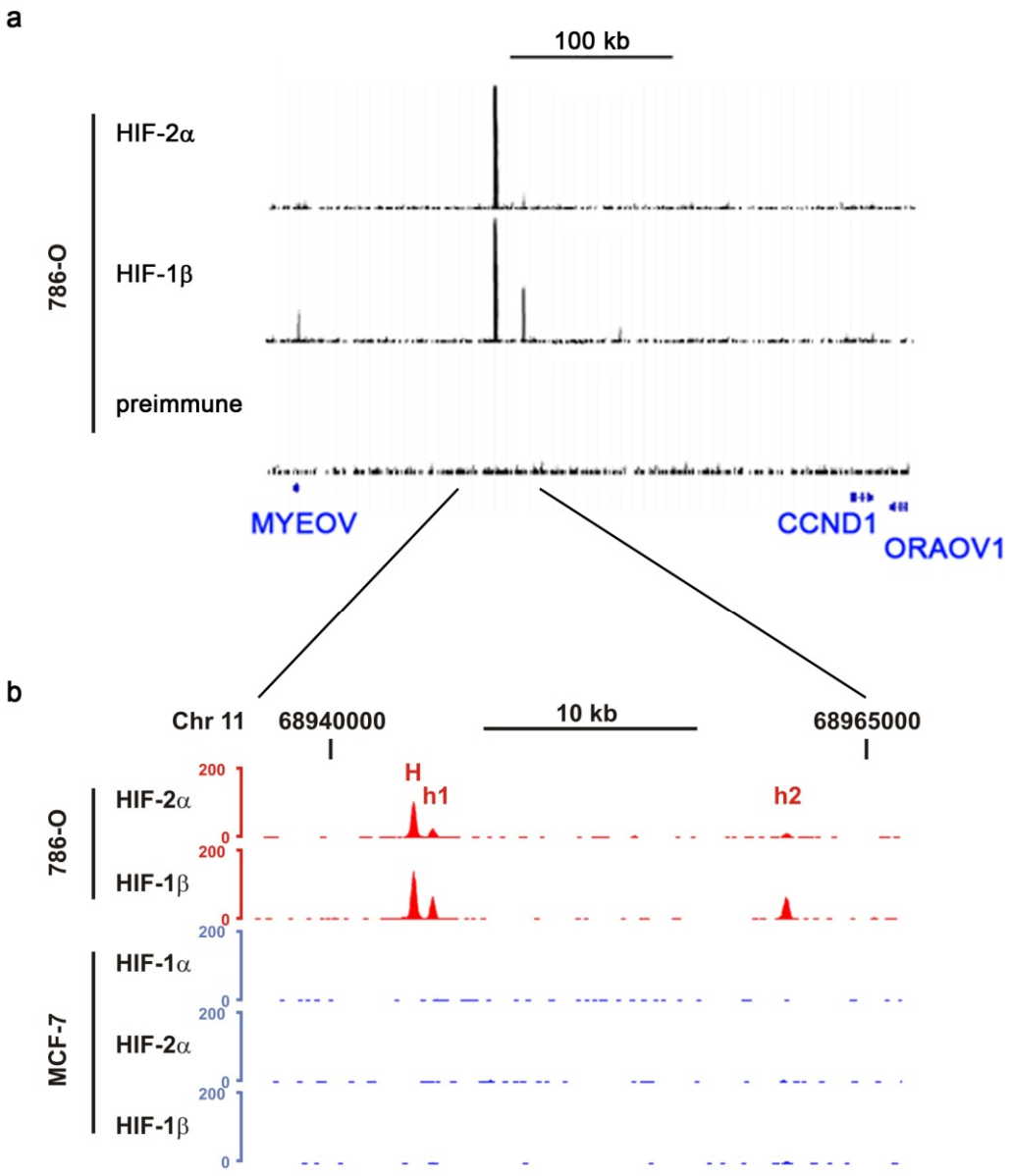


Figure 5.1: (a) Read density tracks from ChIP-seq experiments for HIF-2α,HIF-1β and preimmune control in 786-O cells on 11q13.3 show distinct binding of HIF-2 at a remote locus and relation to the next nearest genes (*MYEOV* centromeric and *CCND1* telomeric). (b) Enlarged view of the binding region reveals a major peak 11q13.3 H and two minor peaks h1 and h2. No HIF-binding is observed in the MCF-7 breast cancer cell line.

To confirm the results from the ChIP-seq studies, I performed ChIP-qPCR experiments against the three HIF-binding sites on 11q13.3 on independent samples from 786-O and MCF-7 cells. In 786-O cells HIF-2 binding to 11q13.3 H and h1 was confirmed whereas no significant HIF-signals for these sites were detected in MCF-7 cells (Figure 5.2). The minor

peak 11q13.3 h2 did not reproduce in ChIP-qPCR in 786-O cells. For this peak only weak signals for HIF-binding were detectable in the ChIP-seq tracks, mainly for HIF-1β. This suggests either that HIF-binding to this locus is very weak and difficult to detect by qPCR or that the signal in the ChIP-seq experiments represents an (false positive) artefact.

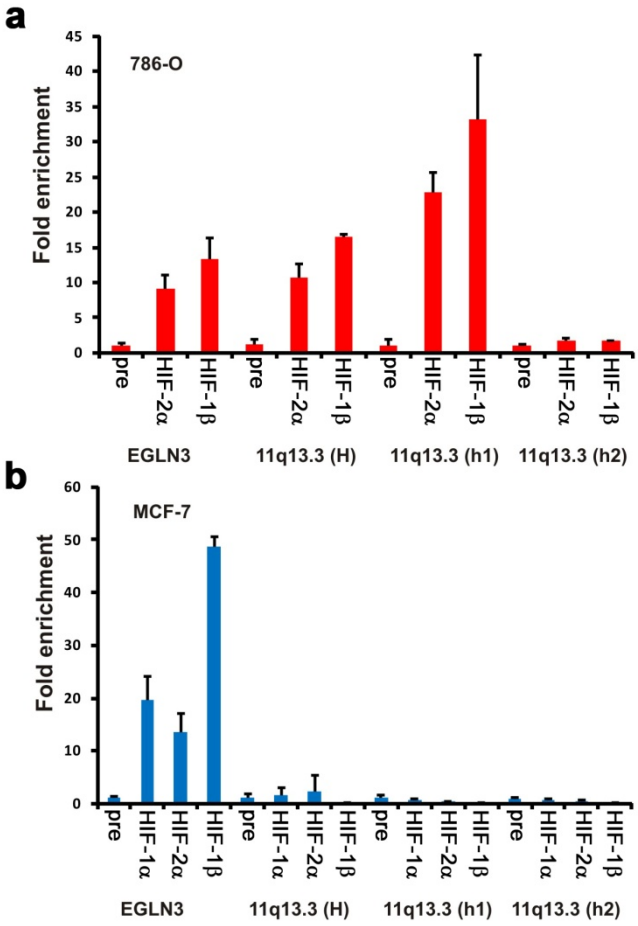


Figure 5.2: ChIP qPCR in independent samples from (a) untreated 786-O cells or (b) MCF-7 cells treated with 1 mM DMOG validating HIF-binding to the three different HIF-binding sites (11q13.3 H, h1 and h2) identified in ChIP-seq experiments in 786-O cells. Enrichment levels for EGLN3 are shown for comparison. Note that only low levels of enrichment were seen at 11q13.3 (h2) which corresponds to a low number of reads at this location in the ChIP-seq experiment. Bars show mean ± s.d. (n=3).

5.2.2 Cell type specific HIF-binding and accessibility

Since HIF-binding at the 11q13.3 locus was observed in 786-O but not MCF-7 cells and regulation of cyclin D1 via loss of VHL is a common feature in renal cancer but no other cell type (Raval, Lau et al. 2005), I next determined whether HIF-binding to this locus was

a specific property of VHL-deficiency that was shared with other ccRCC cell-lines. I therefore examined HIF-binding by ChIP-qPCR at the 11q13.3 HIF-binding sites across a range of pVHL-defective ccRCC cell-lines and a range of cell lines, expressing wild-type VHL (including non cancerous renal epithelial cells) in which HIF had been temporarily stabilised using 1 mM DMOG (Figure 5.3).

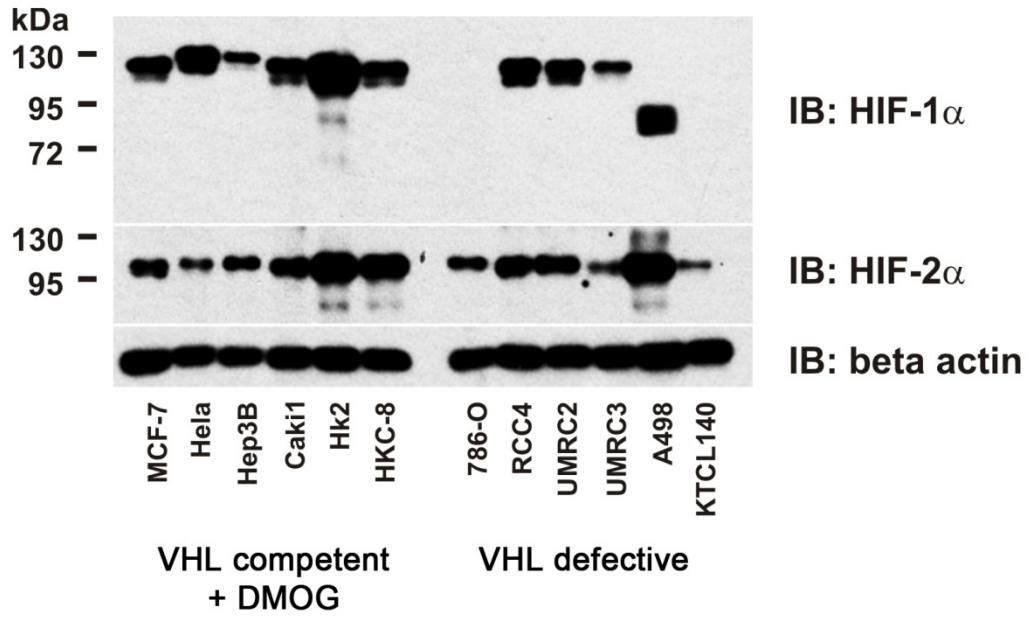


Figure 5.3: Expression of HIF-α isoforms in a panel of pVHL wild-type and pVHL-defective cell lines. Immunoblots (IB) detecting HIF-1α and HIF-2α protein in the panel of six pVHL wild-type cell lines (left) in which HIF was induced using the HIF hydroxylase inhibitor dimethyloxalylglycine (DMOG, 1 mM) and six pVHL-defective cell lines (right). Note that 786-O and KTCL140 have no detectable HIF-1α protein by this assay.

High levels of HIF-2α and HIF-1β binding were seen at the major 11q13.3 H peak in all VHL-defective ccRCC cell-lines with statistically insignificant binding in VHL wild-type cell lines (Figure 5.4), whilst binding was seen in all cell lines at the *EGLN3* locus. These findings imply that HIF is binding to the 11q13.3 locus in a cell-type specific manner and that loss of VHL is a common feature of these cells.

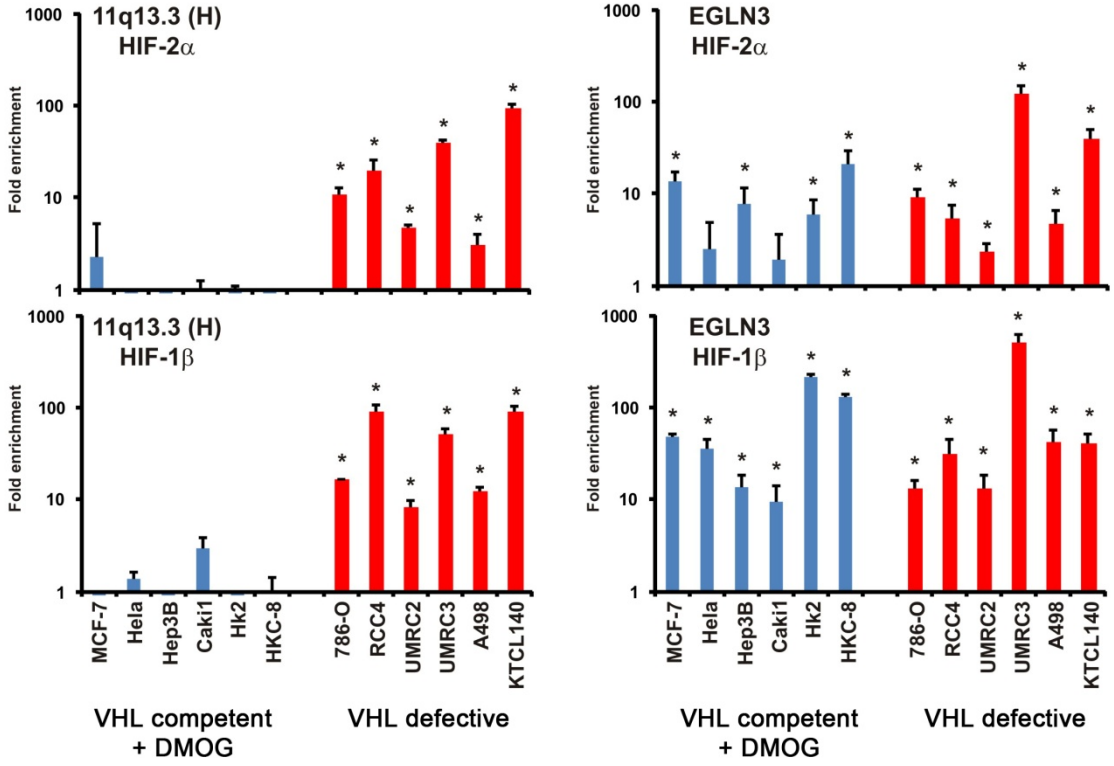


Figure 5.4: ChIP-qPCR results from the same panel of pVHL wild-type (blue) and pVHL-defective (red) cell lines for HIF-2α (upper) and HIF-1β binding (lower) at 11q13.3 (H) (left panel); for comparison signals at the *EGLN3* locus are shown (right panels). Bars show mean ± s.d. (n=3). *) $p < 0.05$ compared to preimmune control.

Previous reports have demonstrated that only HIF-2 (and not HIF-1) is able to regulate cyclin D1 expression in RCCs. If the 11q13.3 HIF-binding locus is a regulatory region for cyclin D1 expression, this raises the question whether HIF-1α is able to bind to this locus at all. Amongst the panel of VHL-defective RCC cell lines which express detectable levels of HIF-1α (Figure 5.3), all showed significant enrichment for this isoform at the 11q13.3 H locus (Figure 5.5). Two VHL-defective RCC cell lines (786-O and KTCL140) do not express HIF-1α and therefore do not exhibit binding of this isoform. All of the VHL wild type cell lines express HIF-1α, but as for HIF-2α and HIF-1β none demonstrated binding at the 11q13.3 HIF-binding site. These results indicate that post-binding mechanisms contribute

to the HIF-isoform specific regulation of cyclin D1 expression following binding at the 11q13.3 locus (as seen for other loci (Lau, Tian et al. 2007)).

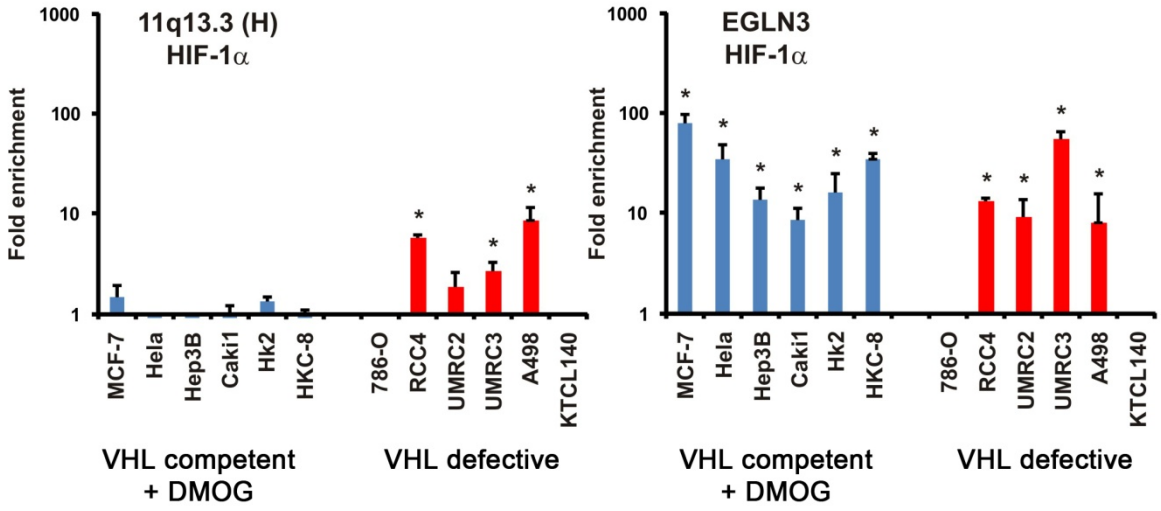


Figure 5.5: HIF-1α ChIP-qPCR in the panel of pVHL wild-type (blue) and pVHL-defective (red) cell lines at 11q13.3 (H) and the control *EGLN3* enhancer site. Data is mean ± s.d. (n=3); *p<0.05 relative to preimmune control.

For completion, HIF-1β binding to the two minor HIF-binding peaks 11q13.3 h1 and h2 was tested. Robust signals were detected in VHL-defective cells for the 11q13.3 h1 whereas 11q13.3 h2 did not reproduce in any of the cell lines tested. These findings indicate that the 11q13.3 H and h1 sites are the main loci for HIF-binding in pVHL-defective renal cancer (Figure 5.6).

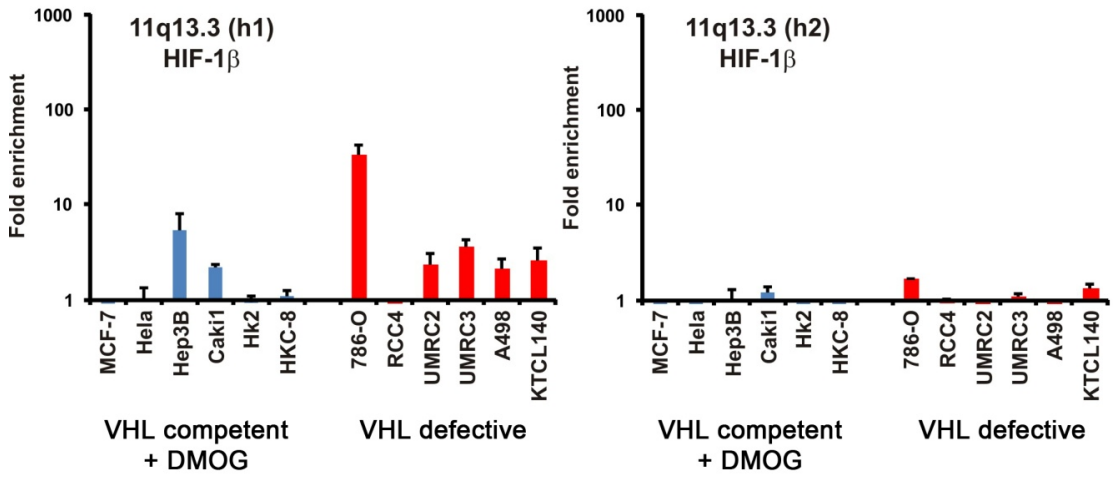


Figure 5.6: HIF-1β ChIP-qPCR in the panel of pVHL wild-type (blue) and pVHL-defective (red) cell lines at the two minor HIF-binding peaks 11q13.3 (h1) and 11q13.3 (h2). Bars show mean ± s.d. (n=3).

As outlined in Chapter 2 and Chapter 4, HIF-binding is associated with DNAse1 hypersensitive sites, which serve as a marker for open chromatin. Cell-type specific binding is associated with cell-specific DNAse hypersensitivity at that site. Although pan-genomic analyses of normoxic DNA accessibility are published for MCF-7 cells, no such analyses are available for 786-O or other renal cancer cells. I therefore examined whether the major HIF-2 binding site H at the 11q13.3 locus exhibited differences in DNA accessibility in pVHL-defective and VHL-competent cells using FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements) analysis. This method enriches for non-protein associated DNA fragments derived from cross linked, sonicated and phenol chloroform extracted chromatin and serves as a correlate for open chromatin (Giresi, Kim et al. 2007). DNA from FAIRE is then quantified by qPCR and compared to input DNA and to DNA from a locus with no open chromatin conformation. Strikingly, as for HIF-binding, a strong correlation between the presence of open chromatin at 11q13.3 H and loss of functional VHL in renal cancer cells was observed across the panel of cell lines, indicating that the 11q13.3 H locus is only accessible in pVHL-defective renal cancer cells (Figure

5.7). FAIRE assays performed for a comparable promoter-remote HIF-binding site at the *NDRG1* gene locus (to which HIF binds in all cell lines (Figure 5.8)) confirmed that the assay had worked in all cell lines tested. These findings implicate that open chromatin at the 11q13.3 locus is also cell-type specific and that it is a prerequisite for HIF-binding to that locus.

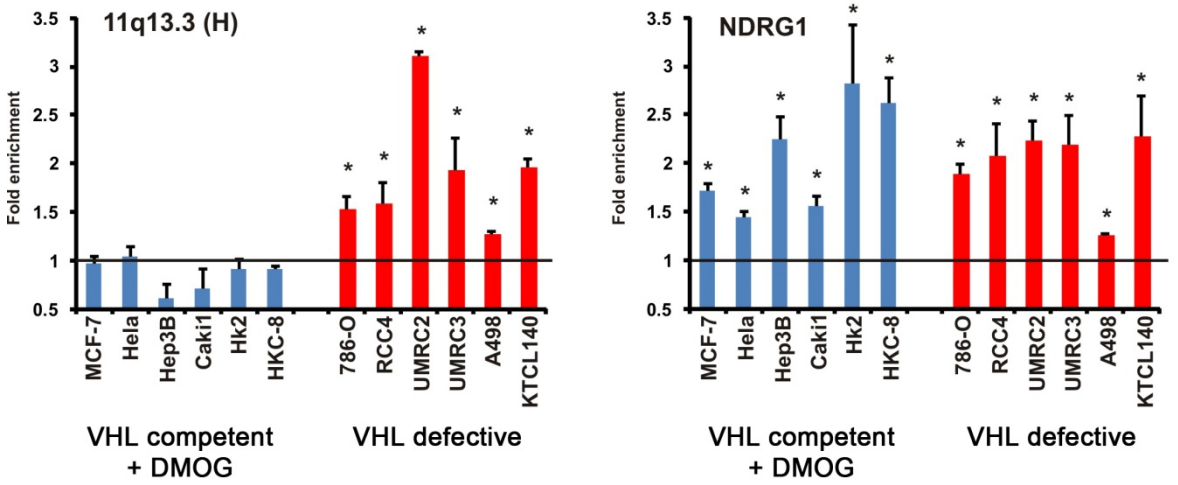


Figure 5.7: Cell-type specificity of the 11q13.3 enhancer. FAIRE analysis is shown for a panel of six pVHL wild-type cell lines (blue) in which HIF was induced using HIF hydroxylase inhibitor dimethyloxalylglycine (DMOG) and six pVHL-defective cell lines (red). qPCR products were normalized as described. Data in the left panel is the enrichment value for primers centred on the ChIP-seq signal (H); for comparison signals at the ubiquitous HIF-binding enhancer (*NDRG1)* are also given (right panel). Bars show mean ± s.d. (n=3). *) p < 0.05 compared to the control site.

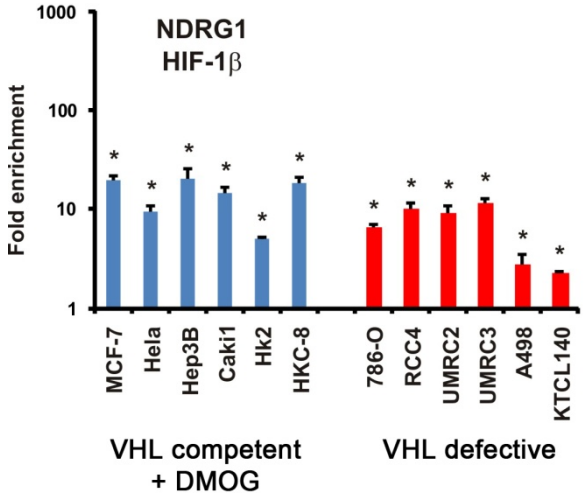


Figure 5.8: HIF-1β ChIP-qPCR in the panel of pVHL wild-type (blue) and pVHL-defective (red) cell lines at the *NDRG1* enhancer site which was used as a control for FAIRE assays. Bars show mean ± s.d. (n=3). *) p < 0.05 compared to preimmune control.

Interestingly, re-expression of wild-type VHL in 786-O cells does not reverse the open chromatin configuration or the ability of these cells to bind HIF at this locus when HIF is temporarily stabilised by DMOG (Figure 5.9). This finding indicates that accessibility at the 11q13.3 locus is a stable feature of pVHL-defective cells which is either acquired and cannot be reversed by VHL reintroduction or is already present in tumour progenitor cells. Importantly, HIF-2α and concurrent overexpression of cyclin D1 can be found in early cystic lesions in the kidneys of patients with VHL Syndrome (Raval, Lau et al. 2005) suggesting that a potential functional link between HIF-2 and cyclin D1 is already present in early tumour stages.

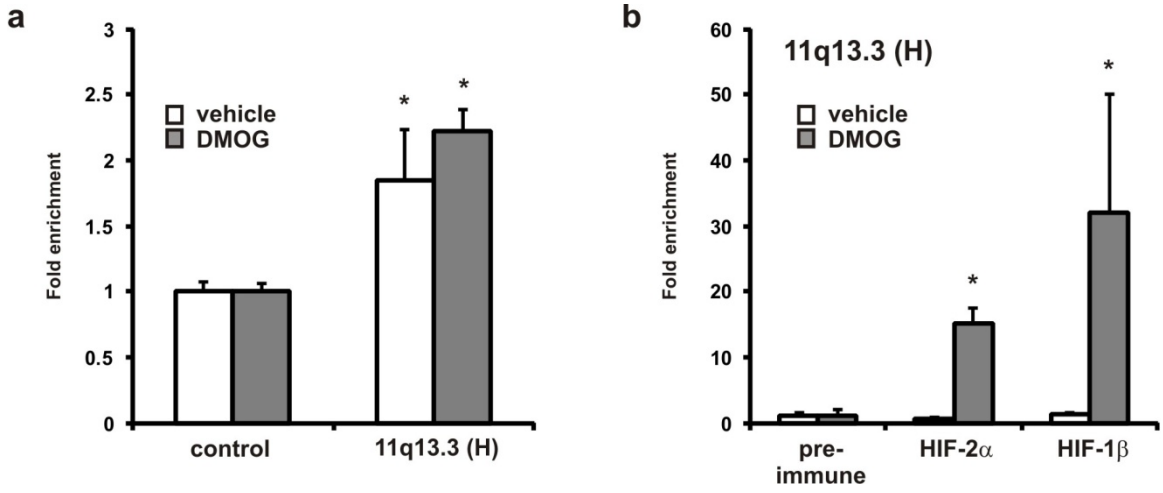


Figure 5.9: VHL re-introduction does not reverse open chromatin configuration at 11q13.3. (a) FAIRE analysis in 786-O cells re-expressing wild type pVHL (786-O VHL cells). Data is the enrichment value for primers centred on the ChIP-seq signal (H) normalized to the control (non-enhancer) site; cells treated with or without DMOG to induce HIF. (b) ChIP-qPCR analysis for HIF-2α and HIF-1β at the 11q13.3 H peak in 786-O VHL cells treated with or without DMOG. Data is mean ± s.d. (n=3); *p<0.05 relative to preimmune (ChIP) or to a control site (FAIRE).

Taken together, HIF-binding at the 11q13.3 locus is specific for pVHL-defective renal cancer cells and, consistent with genome-wide findings, DNA accessibility appears to play a major role in defining this specificity. In addition, binding is not restricted to HIF-2, but – when present – HIF-1 is also able to bind to the 11.q13.3 locus.

5.2.3 The 11q13.3 HIF-binding site exhibits the hallmarks of an active enhancer

To further examine the functional role of the 11q13.3 locus, I decided to examine further epigenetic markers for regulatory regions. As for the ChIP-seq studies I used 786-O and MCF-7 cells to compare pVHL-defective and pVHL-competent cells. In ChIP experiments comparable levels of total histone 3 were detected in the two cell lines at a nearby control region, but reduced levels were measured at the 11q13.3 (H) locus in 786-O cells compared to MCF-7 cells indicating a reduced occupancy of this locus by histones in 786-O cells (Figure 5.10). To further fine map the region of open chromatin conformation at the 11q13.3 locus I performed a tiled qPCR FAIRE analysis across the HIF-binding site was conducted (Figure 5.11a). When compared to a nearby control region enrichment of open chromatin was seen across a region of about 2 kb covering the HIF-binding site in 786-O cells but not in MCF-7 cells. Again, findings for this site are consistent with the model in which cell-specific HIF-binding correlates with cell-specific DNA-accessibility (Chapter 4).

Next, to determine whether this binding site has the features of a functional enhancer and to exclude the existence of a cryptic promoter, the locus was tested for the epigenetic features of both promoters and enhancers. High levels of H3K4me1 and H3K27ac, together with low levels of H3K4me3 histone modifications were observed by ChIP-qPCR at the major HIF-binding peak 11q13.3 H indicating that this site has the epigenetic hallmarks of an active enhancer and excluding the possibility that it acts as a cryptic promoter (Figure 5.11b-d) (Heintzman, Stuart et al. 2007; Creyghton, Cheng et al. 2010; Rada-Iglesias, Bajpai et al. 2011). Furthermore, in concordance with the findings for the remote HIF-binding sites *ARRDC3* and *ERRFI1* in Chapter 2 low level interaction between the HIF-binding site and RNA pol2 was observed in 786-O cells, but not in MCF-7

cells (Figure 5.11e) indicating a cell-specific interaction with the basal transcriptional machinery (Heintzman, Stuart et al. 2007). Finally, the main HIF-binding site 11q13.3 H contains two consensus HIF-binding motifs one at the centre of the ChIP-seq peak and one close to the periphery (Appendix 8.5). Sequences spanning the major HIF-binding site 11q13.3 H and containing these HREs were able to enhance expression of a luciferase reporter gene driven by the SV40 promoter when transfected into 786-O cells in a manner that was dependent on an intact central HRE (Figure 5.11f). Sequences containing both the major peak 11q13.3 H and the minor peak 11q13.3 h1 (which does not contain an HRE, Appendix 8.5) showed still greater enhancing activity, but were only partially inhibited by mutating the central HRE in the major peak 11q13.3 H (Figure 5.11f). Taken together, these data indicate that the 11q13.3 H site is acting as a HIF-dependent neo-enhancer in 786-O cells.

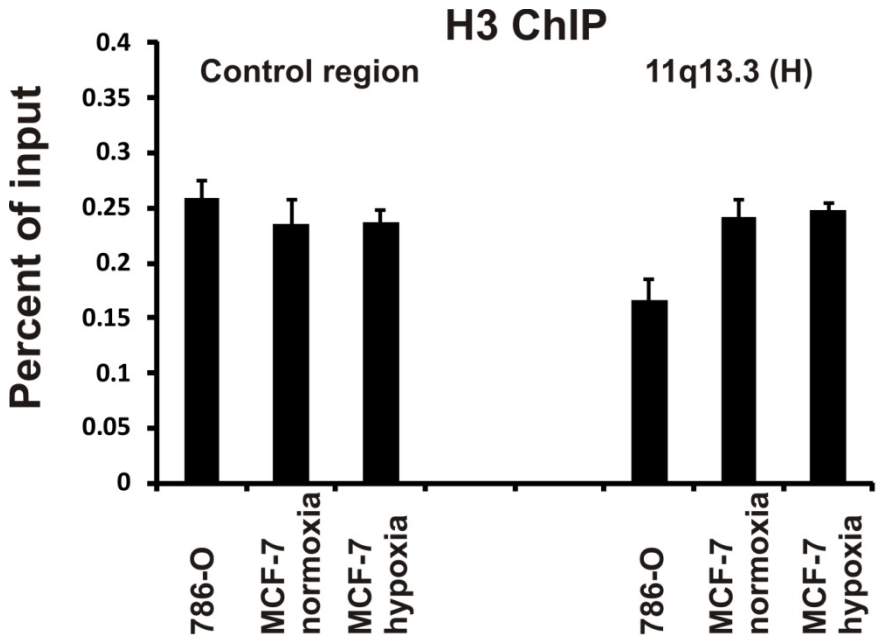


Figure 5.10: Total histone 3 (H3) levels at a control region (CCND1 enhancer negative control) and the 11q13.3 (H) HIF-binding site. ChIP experiments against H3 reveal comparable levels of H3 at the control region (putative non open chromatin) and reduced levels in 786-O cells at the 11q13.3 enhancer. Bars show mean ± s.d. (n=3)

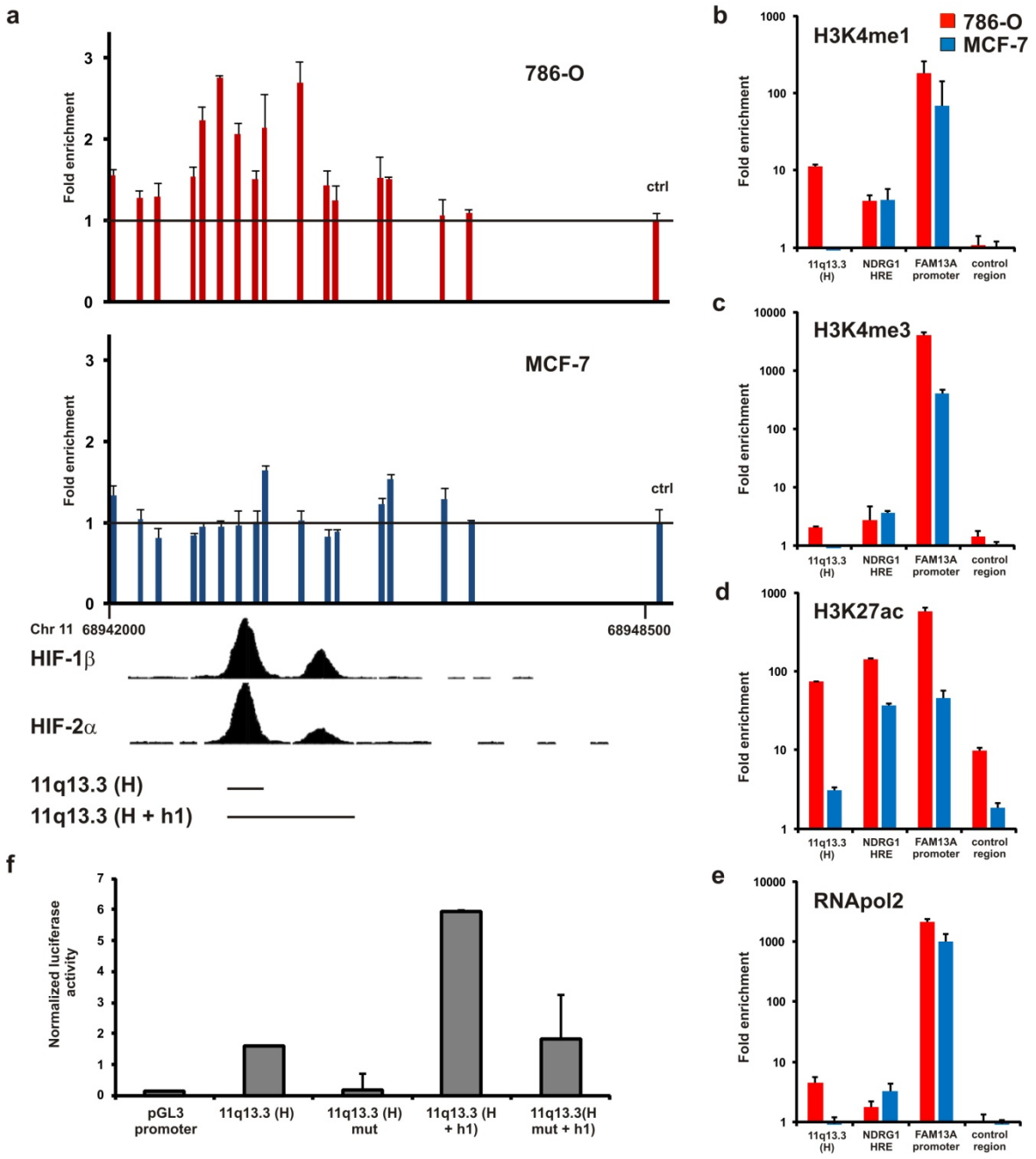


Figure 5.11: Chromatin structure and function of an enhancer at 11q13.3. (a) Formaldehyde-assisted isolation of regulatory elements (FAIRE) was performed in 786-O and MCF-7 cells to map nucleosome occupancy at the 11q13.3 locus. qPCR results were normalized to those obtained for input DNA at this locus and then related to values obtained at a nearby control region outside the putative enhancer (ctrl, far right bar). Data is plotted against chromosomal location; the position of the ChIP-seq signals for HIF-1β and HIF-2α is illustrated below. (b-e) ChIP-qPCR signals at 11q13.3 (H) obtained from 786-O cells (red) or MCF-7 cells (blue) using antibodies against (b) H3K4me1, (c) H3K4me3, (d) H3K27ac and (e) RNA pol2; for comparison signals at an ubiquitous HIF-binding enhancer (*NDRG1)*, a HIF-binding promoter (*FAM13A1)* and a nearby non-enhancer control region are shown. Bars in FAIRE and ChIP experiments show mean ± s.d. (n=3). (f) Luciferase reporter assay performed in 786-O cells using the sequences 11q13.3 (H) or 11q13.3 (H+h1) (indicated above); mut indicates mutation of the core HRE motif in H. Data is normalized to co-transfected β-galactosidase activity, mean ± SD (n=3).

5.2.4 The 11q13.3 H locus is the main HIF-binding regulatory site

Several lines of evidence indicate that the major site, 11q13.3 H, is of predominant importance for HIF-binding in renal cancer. Firstly, out of the three sites that show HIF-binding signals in ChIP-seq experiments it is the only locus reaching high-stringency levels of HIF-binding on a genome-wide level (indeed it is ranked 12th out of the 608 high-stringency sites). Secondly, this site contains two HREs, one of which is necessary to induce enhancer activity in reporter assays (see above). The minor site 11q13.3 h1 does not contain a HRE, whereas 11q13.3 h2 contains two HREs (Appendix 8.5), but a sequence covering this locus alone, does not enhance reporter activity of the pGL3 promoter vector in transfected 786-O cells (Figure 5.12). Thirdly, ChIP-qPCR validation shows that in independent samples from 786-O cells the 11q13.3 h2 locus only enriches HIF-binding at very low levels compared to the other two loci (Figure 5.2) and that in the panel of pVHL-defective cell lines no other cell line apart from 786-O cells showed significant HIF-1β binding at 11q13.3 h2 site (Figure 5.6).

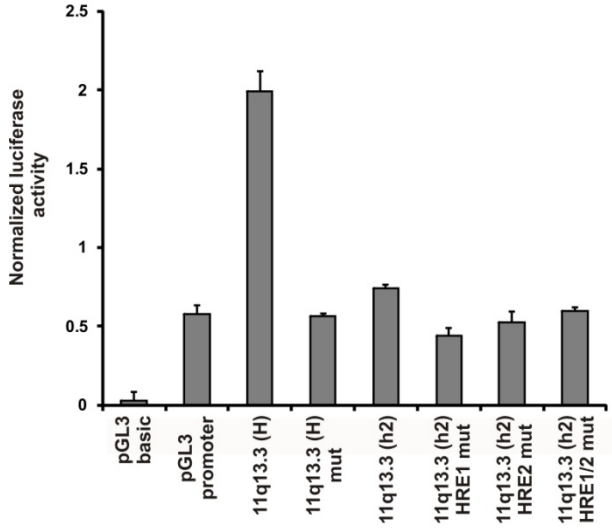


Figure 5.12: Reporter assay testing the ability of 11q13.3 (h2) to enhance luciferase activity. As outlined above a 227 bp sequence covering the HRE of 11q13.3 (H) is able to increase reporter activity. No significant induction is seen when testing a sequence covering the minor peak 11q13.3 (h2). Two HREs present in this sequence were mutated individually or combined (HRE1mut, HRE2mut, HRE1/2mut) with no effects on reporter activity. Bars show mean ± s.d. (n=3).

5.2.5 HIF regulates cyclin D1 and the enhancer loops to the *CCND1* promoter

Expression analyses in renal cancer suggest that cyclin D1 expression is regulated via HIF-2 in renal cancer. The existence of a functional enhancer specific to pVHL-defective renal cancer cells about 220 kb upstream of the *CCND1* gene implies that this is the regulatory element conferring HIF-dependent regulation of cyclin D1 expression. To determine which genes are regulated via this newly identified enhancer the expression arrays introduced in chapter 4 were interrogated to identify neighbouring genes that are regulated by HIF or by hypoxia and which are deregulated in renal cancer. In all arrays *CCND1* was the most regulated gene within 1Mb of 11q13.3 H (Figure 5.13). Regulation of cyclin D1 expression by hypoxia or HIF2-α was confirmed by qPCR in 786-O cells (Figure 5.14). These findings strongly imply that HIF-binding to the 11q13.3 H enhancer is indeed modulating the expression of cyclin D1.

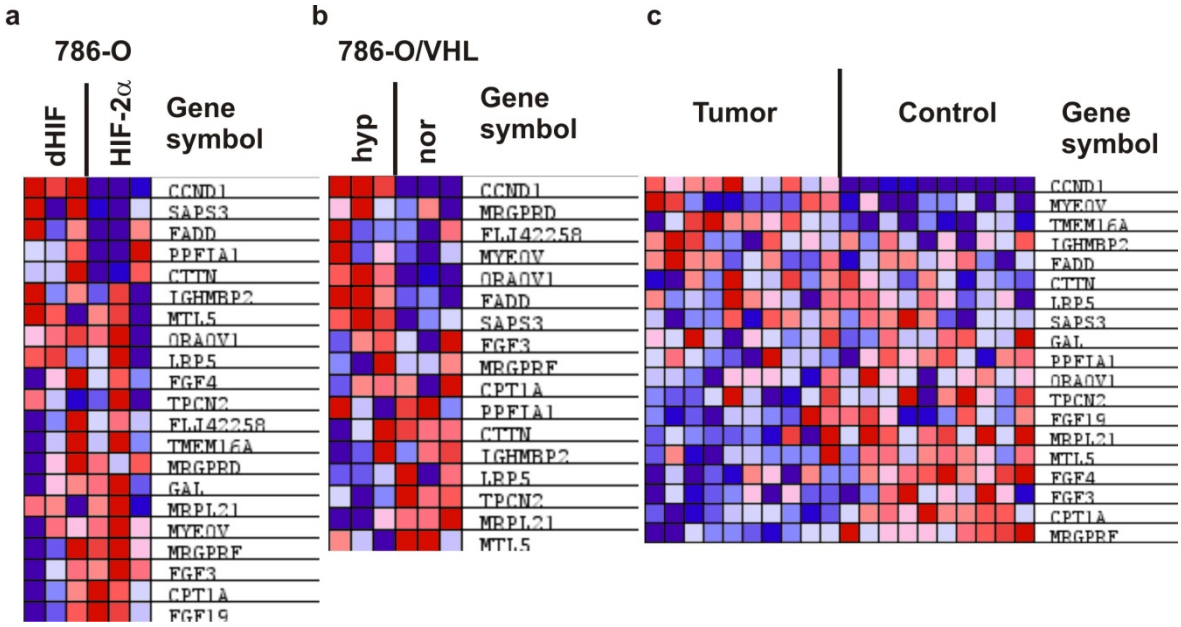


Figure 5.13: Heat maps representing microarray analyses of gene expression within 1Mb of the 11q13.3 HIF-binding site. Each square represents a normalised expression level of the respective gene in one sample. The darker the red colour the higher is the expression level. The darker the blue colour the lower. (a) Relative levels of expressed genes in 786-O cells treated with control siRNA (drosophila HIF, dHIF) or siRNA targeting HIF-2α (b) Relative levels of expressed genes in 786-O/VHL cells re-expressing pVHL exposed to normoxia (nor) or hypoxia (hyp, 1% O_2) for 16 h. (c) Relative expression levels in normal kidney samples and renal tumours; data derived from Gumz et al. ((Gumz, Zou et al. 2007) GSE6344).

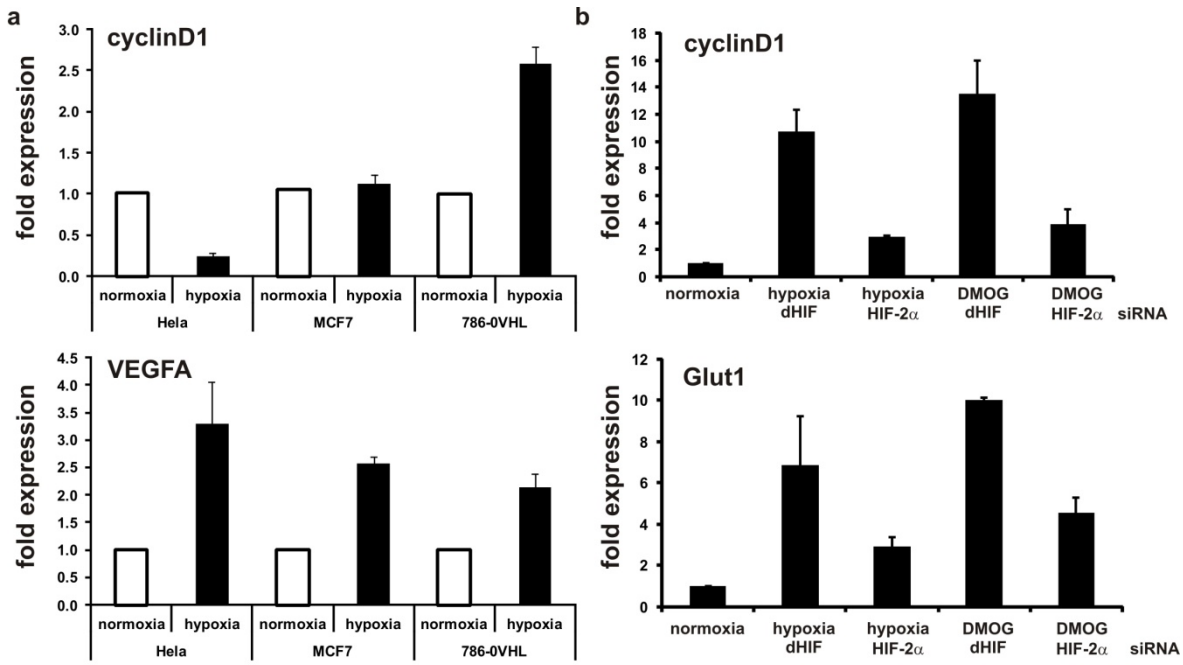


Figure 5.14: (a) Expression qPCR analysis of cyclin D1 and VEGFA in Hela, MCF-7 and 786-O VHL cells exposed to normoxia or 1% O_2 for 16 hours. Only in 786-O VHL cells cyclin D1 expression is induced by hypoxia. (b) Expression qPCR analysis of cyclin D1 and Glut1 mRNA in 786-O VHL cells treated with hypoxia or DMOG and transfected with either control siRNA (dHIF) or siRNA directed against HIF-2α as indicated.

Long-range gene regulation requires interaction of the distant regulatory element with promoters of the regulated genes. To test the hypothesis that the distant HIF-binding site on 11q13.3 is regulating cyclin D1 expression more formally, additional experiments were performed to determine the presence of a physical association between this HIF-binding site and the *CCND1* promoter. Veronica Buckle and Jill Brown (WIMM, University of Oxford, U.K.) performed high-resolution 2D and 3D fluorescent in situ hybridisation (FISH) experiments using fluorescent labelled FISH probes at the HIF-binding site, the *CCND1* promoter and control sites equidistant in both the centromeric and the telomeric direction. A high resolution imaging technique and computer-assisted distance measurement were used to calculate interprobe distances from about 100 nuclei per fosmid combination. Results from 2D and independent 3D experiments showed that the

HIF-binding site and the *CCND1* promoter were more closely associated with each other, than with probes of comparable genomic-distances in opposing directions, in 786-O cells, but not in MCF-7 cells (Figures 5.15 and 5.16).

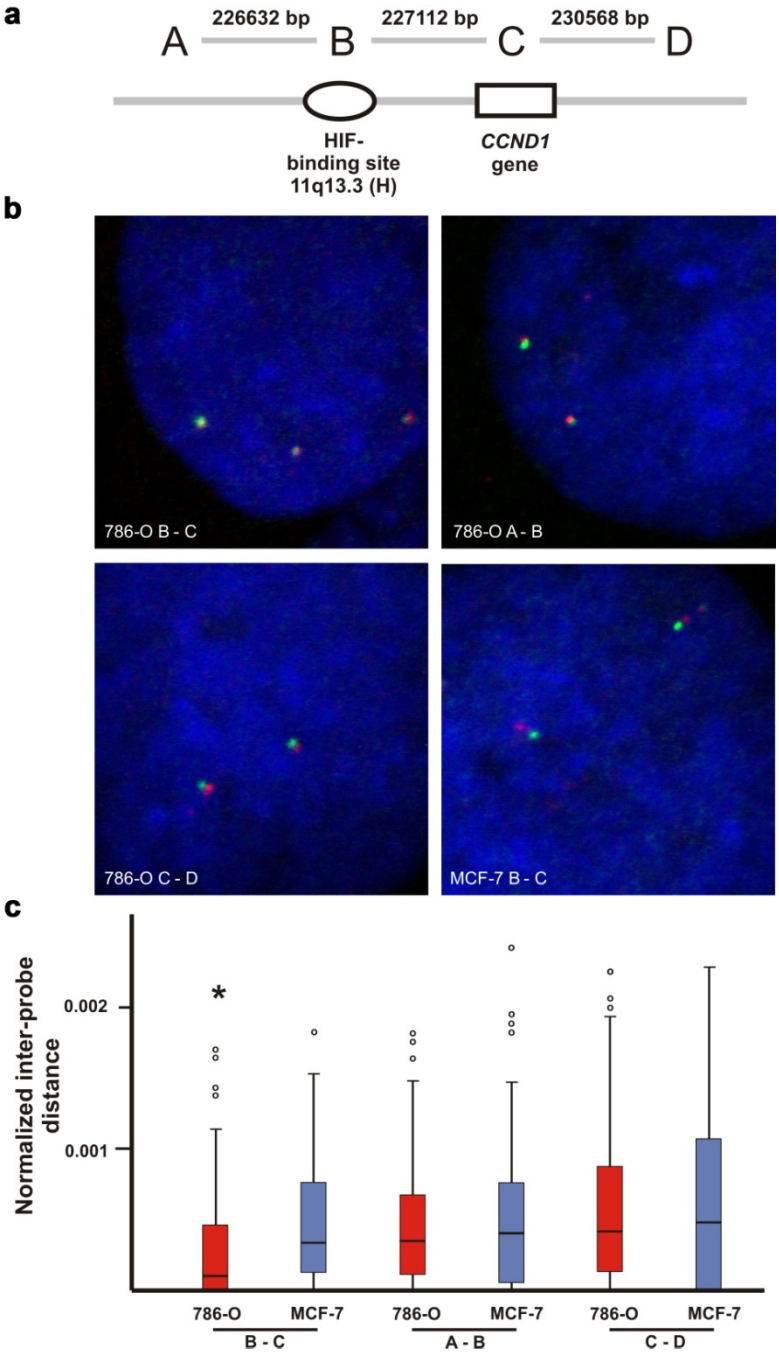


Figure 5.15: 2D FISH (a) Location and distance between midpoints of fosmid probes used for 2D fluorescent *in situ* hybridization (FISH). (b) Representative images of 2D FISH in 786-O and MCF-7 cells exemplifying signals from the indicated pair of probes (red and green). (c) Box-and-whisker plots showing median, inter-quartile range and full range of visualized inter-probe distances for each pair of probes, normalised to nuclear diameter. Kruskal-Wallis 1-way ANOVA was significant (P<0.001). Mann-Whitney U-test showed that probes at the HIF-binding site and the *CCND1* gene were significantly (*, p<0.001) more closely associated (in 786-O, but not MCF-7 cells) than were other probe pairs.

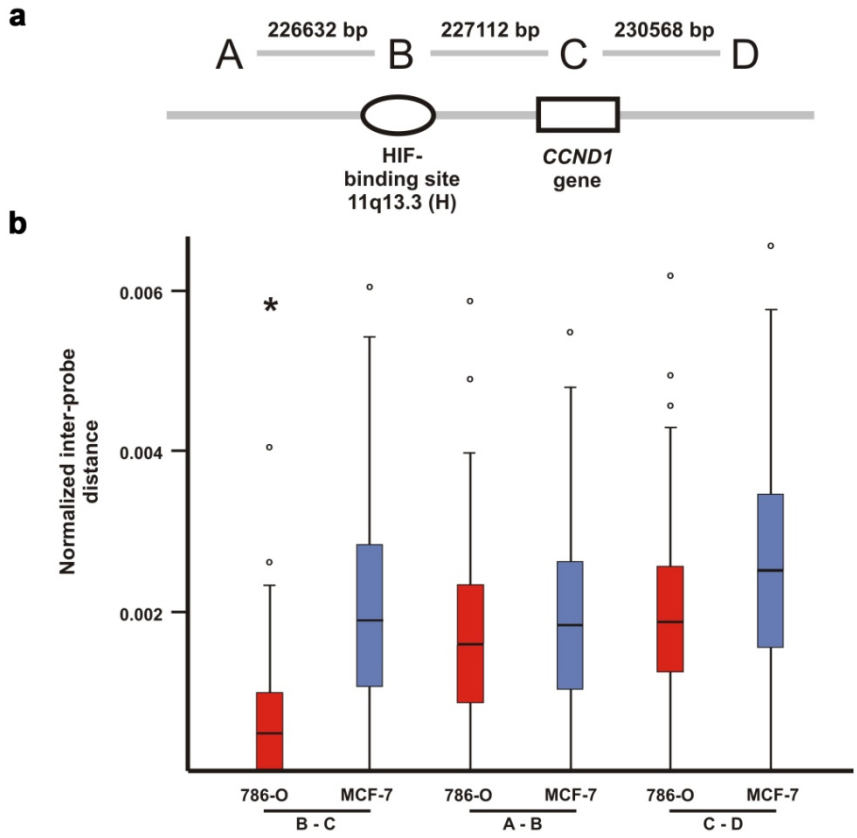


Figure 5.16: 3D FISH. Physical association of the remote HIF-binding site with the *CCND1* promoter. (a) Location and distance between midpoints of fosmid probes used for 3-dimensional (3D) fluorescent *in situ* hybridization (FISH). (b) Box-and-whisker plots showing median, inter-quartile range and full range of visualized inter-probe distances for each pair of probes, normalised to nuclear volume. Kruskal-Wallis 1-way ANOVA was significant (P<0.001). Mann-Whitney U-test showed that probes at the HIF-binding site and the *CCND1* gene were significantly (*, p<0.001) more closely associated (in 786-O, but not MCF-7 cells) than were other probe pairs.

Although the spatial association demonstrated by the FISH assays is strongly suggestive of a direct interaction between enhancer and promoter, an actual physical interaction between enhancer and promoter cannot be proven by this method. Therefore, I used Chromatin Conformation Capture (3C) assays to test for a direct physical interaction between the enhancer and the promoter. In this assay EcoRI restriction enzyme digest of formaldehyde cross linked chromatin was performed. EcoRI restriction sites are distributed across the genome in such a frequency that restriction events occur every few thousand base pairs. For the 11q13.3 locus approximately 50 EcoRI restriction sites are located between putative enhancer and *CCND1* promoter. Restriction enzyme digestion

efficiency was monitored using primer pairs spanning EcoRI sites. Restriction efficiency of >60% was achieved which is of the required order of magnitude suggested by the existing literature (Hagege, Klous et al. 2007). Besides stabilising DNA-protein interactions, cross linking with formaldehyde should also bridge protein-protein interactions thereby preserving DNA-looping interactions over long genomic distances. Subsequent dilution of the restriction digest removes intervening non-cross linked DNA fragments and in a re-ligation assay ligation of DNA fragments in close physical proximity are favoured (see Figure M6 in Methods). After de-cross-linking and DNA isolation newly ligated fragments can be detected using primers located on each side of the newly formed re-ligation site. In the 3C assays presented in this work I have chosen the EcoRI restriction site telomeric and closest to the HIF-binding enhancer as the "anchor". Combinations of a forward primer and a fluorigenic probe close to this site with a series of reverse primers located at EcoRI restriction sites close to the enhancer, in intervening regions or close to the *CCND1* promoter were used for qPCR analysis. This allows calculation of the interaction frequency of each fragment tested with the 11q13.3 enhancer.

BAC clones of this region were digested with EcoRI and re-ligated to generate a library of newly ligated fragments that served as normalisation controls in qPCR assays. To compare results between cell lines interaction frequencies of ubiquitous interacting fragments at the EEF1G locus were determined and used to normalise data (Ramachandrareddy, Bouska et al. 2010). Using this method results showed that the 11q13.3 H HIF-binding site preferentially ligated with sites at the *CCND1* promoter in 786-O and UMRC3 renal cancer cells, but not in MCF-7 cells (Figure 5.17). This implies that there is indeed a physical interaction between the 11q13.3 enhancer and the *CCND1* promoter in renal cancer cell.

Taken together, the findings describe a HIF-binding neo-enhancer of cyclin D1 expression in renal cancer cells.

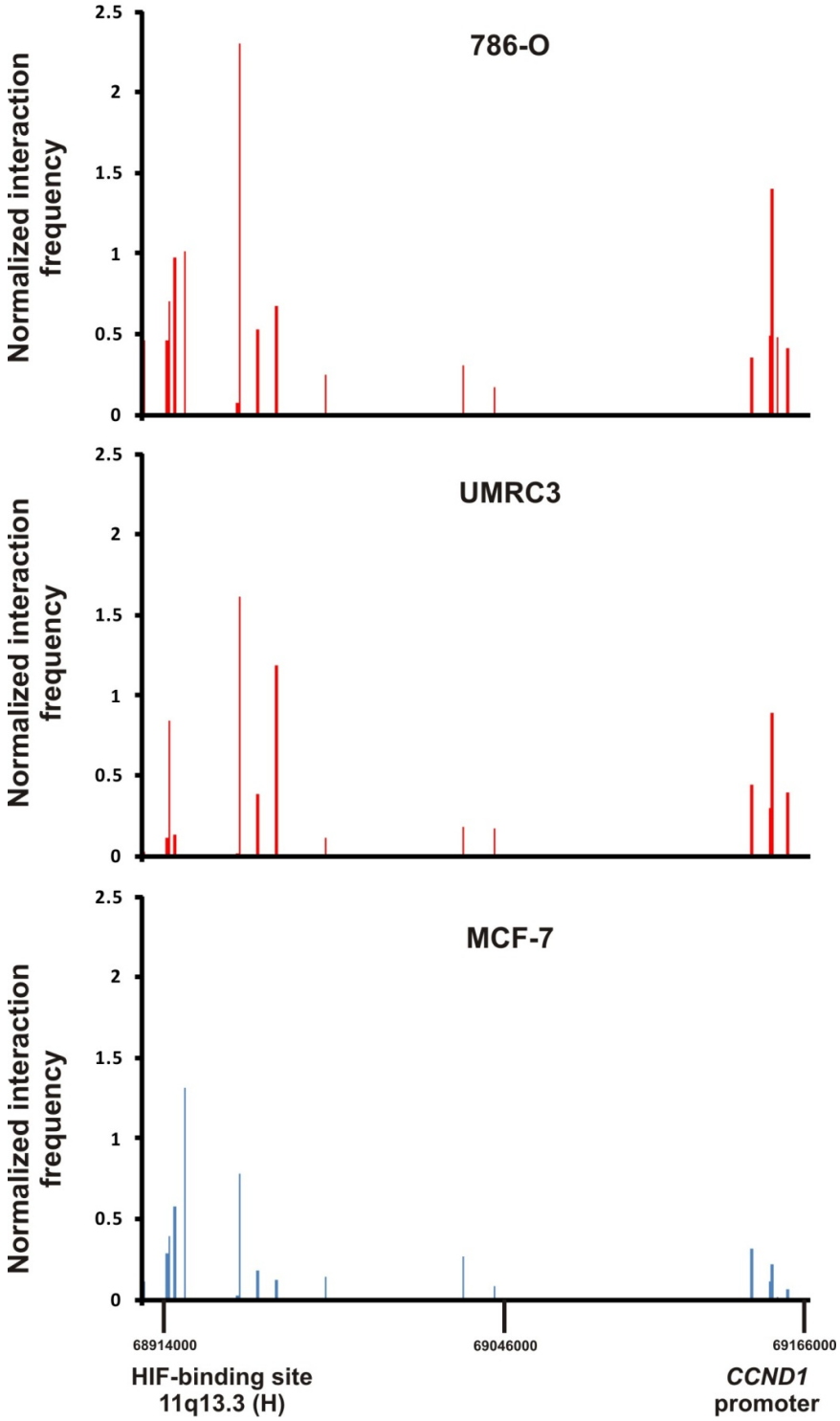


Figure 5.17: Chromatin conformation capture (3C) assays using the remote enhancer site 11q13.3 as the anchor. Bars show qPCR analysis of the re-ligation frequency following restriction enzyme digestion, for cellular DNA, normalized to that of the BAC clones spanning the region of interest. Sites at the 11q13.3 HIF-binding locus preferentially ligated with fragments at the *CCND1* promoter in 786-O and UMRC3 cells, but not in MCF-7 cells indicating a cell-type specific physical interaction. The horizontal axis shows genomic coordinates of chromosome 11.

5.2.6 Genetic evidence that the enhancer is relevant in renal cancer

Purdue et al found three SNPs, which on a genome-wide basis were significantly associated with susceptibility to renal cancer, but no evidence concerning functional effects of polymorphisms at these loci was provided (Purdue, Johansson et al. 2011). Two of these SNPs are located in the first intron of the *EPAS1* gene, encoding for HIF-2α reinforcing the importance of this gene for RCCs. Interestingly, the third SNP rs7105934 was mapped to an intergenic region on 11q13.3 which coincides with the region identified in this work as being a VHL-specific enhancer region for cyclin D1 expression. Closer inspection of the relation between the SNP rs7105934 and the HIF-binding site revealed that the SNP is located approximately 5 kb downstream of the enhancer (Figure 5.18). A thorough genetic analysis of this locus was then performed by Ian Tomlinson and Chiara Bardella (Wellcome Trust Centre for Human Genetics, University of Oxford, UK). First, to test whether there is a genetic link between the tagSNP rs7105934 and the HIF-binding sites, SNPs in the HIF-binding loci were tested for linkage disequilibrium with the tagSNP rs7105934. Analyses of existing databases from the 1000 Genomes Pilot 1 data suggested that there is high linkage disequilibrium (r^2=1) between SNPs at these two sites. To confirm this, DNA from 192 cancer free individuals was genotyped for each SNP and the pairwise linkage disequilibrium (LD) was determined. Figure 5.18 shows the genomic locations of the HIF-binding sites in relation to the SNPs tested for LD. Results confirmed that there is high LD between the tagSNP rs7105934 and the SNPs in the binding regions of HIF. Because not all SNPs in this region were covered by the GWAS analysis, the tagSNP rs7105934 can serve as a proxy for other, potentially functional SNP(s), which could be located anywhere within the region of high LD. Interestingly,

rs4980785 is located just 200 bp upstream of the HIF-binding site 11q13.3 H, but LD with the tagSNP is low indicating that this SNP is not associated with disease.

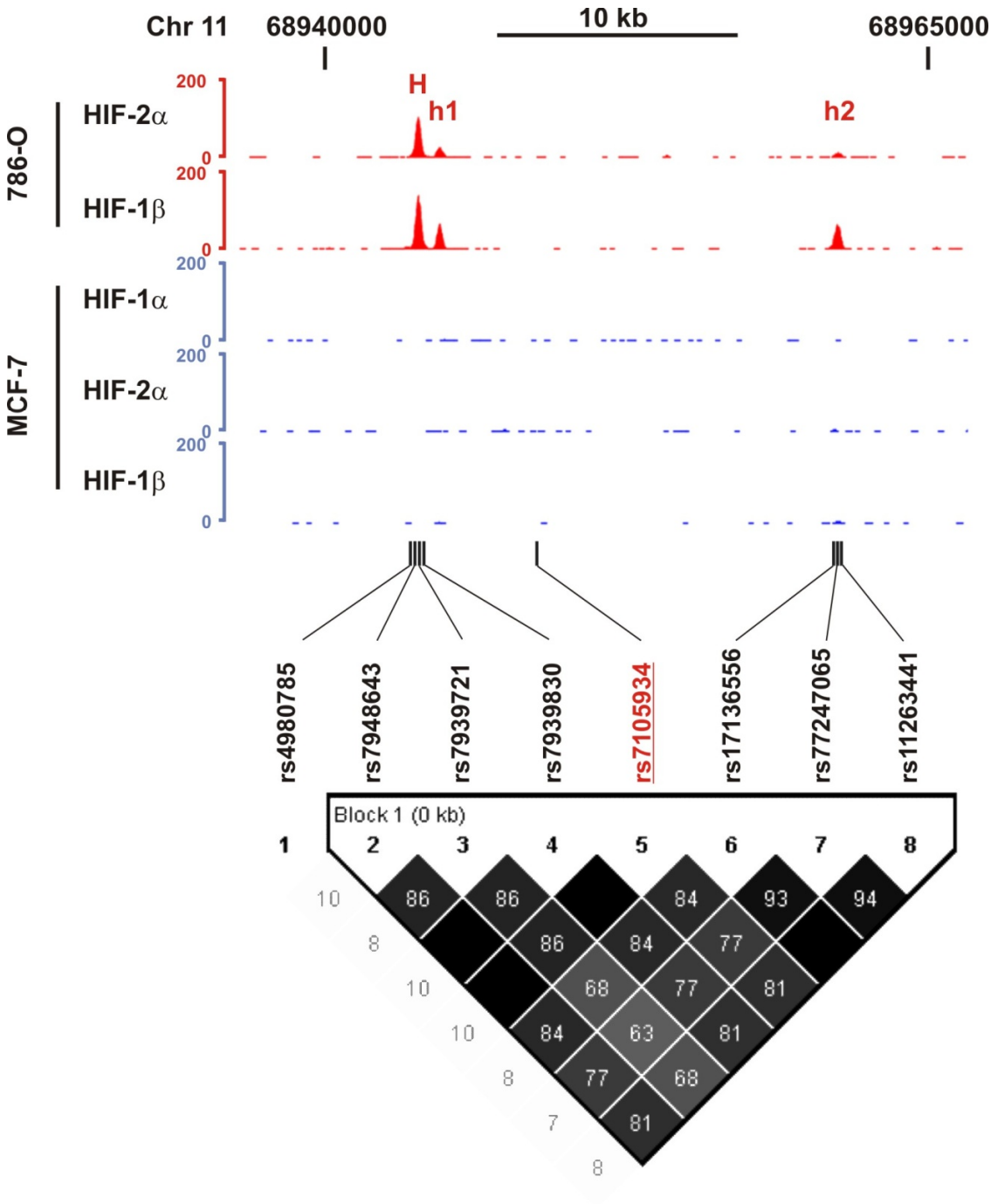


Figure 5.18: HIF-binding at 11q13.3. (a) Genome Browser tracks showing the read density at the 11q13.3 locus for HIF-2α and HIF-1β ChIP-seq in 786-O cells and for HIF-1α, HIF-2α and HIF-1β ChIP-seq in MCF-7 cells in which HIF-α chains were stabilised by hydroxylase inhibition. Three HIF-2- binding sites (H, h1 and h2) are identified in 786-O but not MCF-7 cells. Also shown are the positions of genotyped SNPs, together with LD analysis (r^2 values). SNPs overlapping peak H and peak h2 are in sufficiently strong LD with the disease-associated tagSNP, rs7105934 (red) that the latter can act as a simple proxy.

In summary, the results from genetic analysis show that the region of high linkage disequilibrium at the 11q13.3 renal cancer susceptibility locus extends to the 11q13.3 (H) HIF-binding site. HIF-binding to this site is highly reproducible in pVHL-defective cell lines; it contains a functional HRE and is enhancing cyclin D1 expression. In the next section I will examine whether genetic variants at this site are influencing binding and transcriptional activity of HIF.

5.2.7 11q13.3 polymorphisms influence HIF-binding to the *CCND1* enhancer

The importance of VHL loss and HIF activation to the pathogenesis of ccRCC, the early and ubiquitous development of this site as a neo-enhancer for cyclin D1 expression, the central role of cyclin D1 in control of the cell-cycle and the proximity of this site to RCC associated SNPs lead me to the hypothesis that the rare allele that provides protection against the development of ccRCC may disrupt the operation of the enhancer and/or its ability to bind HIF. Importantly, this protective haplotype extends to and includes 2 SNPs (rs7948643 T>C and rs7939721 C>T) that lie within the main HIF-binding site 11q13.3 H and closely flank the central HRE (Figure 5.19), but are not predicted to directly affect the integrity of the core HRE motif. Mutation of these SNPs to the minor allele in luciferase reporter assays reduced activity to levels comparable to vectors containing either the mutant HRE or no insert at all. Interestingly, introduction of each SNP individually did not cause significant reduction in reporter activity (Figure 5.19) pointing to a potential role for both SNPs in combination in regulating HIF activity. However, reporter assays are highly artificial and do not reflect the *in vivo* situation with chromatinised DNA. Moreover, *in vitro* HIF-binding studies such as electromobility shift assays and DNAse footprinting that could test the effect of individual SNPs on HIF-binding directly suffer from the same limitations. The current work prioritised to screen for effects in an *in vivo* situation first

and therefore further cell-based assays were conducted. Since, the high linkage disequilibrium and close physical proximity (35 bp) of the two SNPs render it extremely unlikely that either would be inherited independently of the other these studies examined the effect of the entire haplotype rather than individual SNPs.

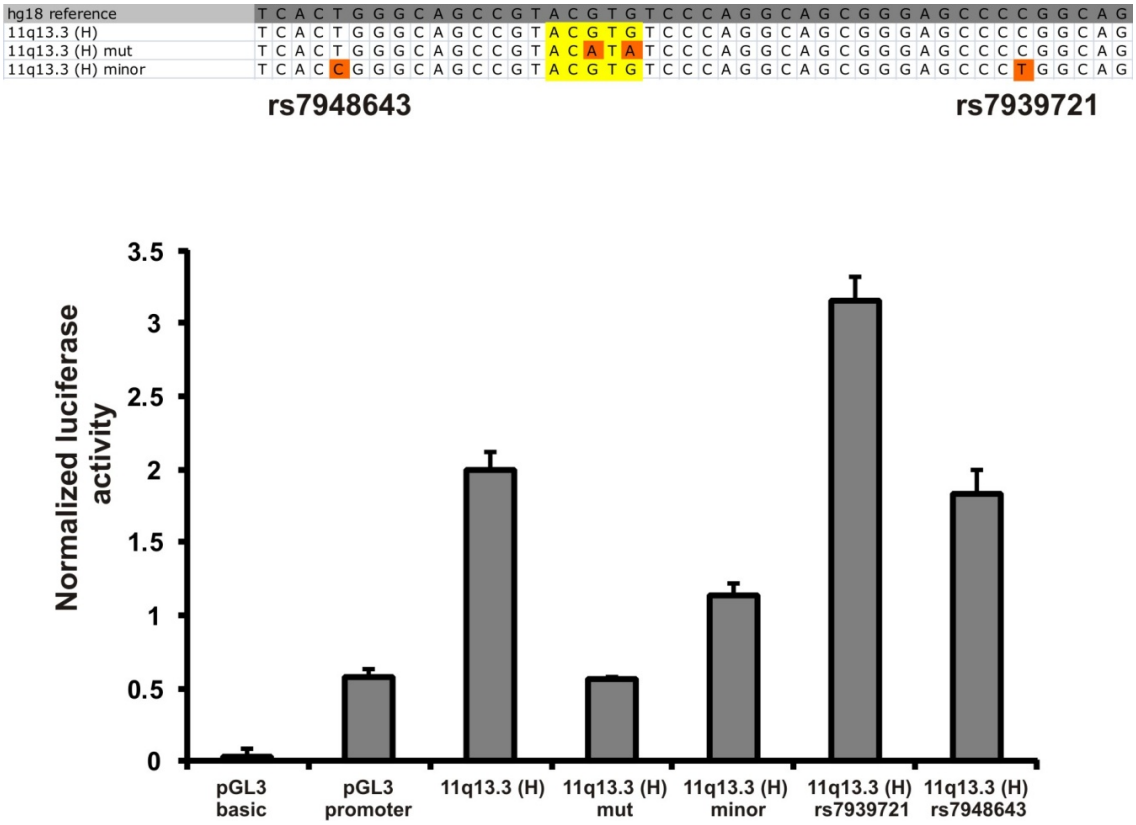


Figure 5.19: The DNA sequence of the 11q13.3 (H) HIF-binding site with the nearest reported SNPs rs7948643 and rs7939721 is shown. The core HRE is highlighted in yellow. For reporter assays the HRE was mutated as indicated in the 227 bp insert used in Figure 5.11. The same 227 bp sequence covering the minor allele was amplified using genomic DNA from an individual homozygous for this allele. To test effects of individual SNPs site directed mutagenesis was carried out to introduce single base pair mutations generating the rs7939721 or rs7948643 SNPs.

To determine whether the rare protective haplotype might affect HIF-binding to the *CCND1* enhancer *in vivo* a number of pVHL-defective ccRCC cell lines were genotyped at this locus (rs7105934, rs7948643 and rs77247065) using allele specific Taqman® PCR assays to screen for heterozygosity. One of the cell lines (KTCL140) was heterozygous for the common and rare alleles. Results from a SNP array using this cell line showed a normal copy number for the 11q13.3 region relative to the remainder of the genome

(Figure 5.20). All genotyping experiments were carried out by Chiara Bardella and Ian Tomlinson (WTCHG, University Oxford, U.K.).

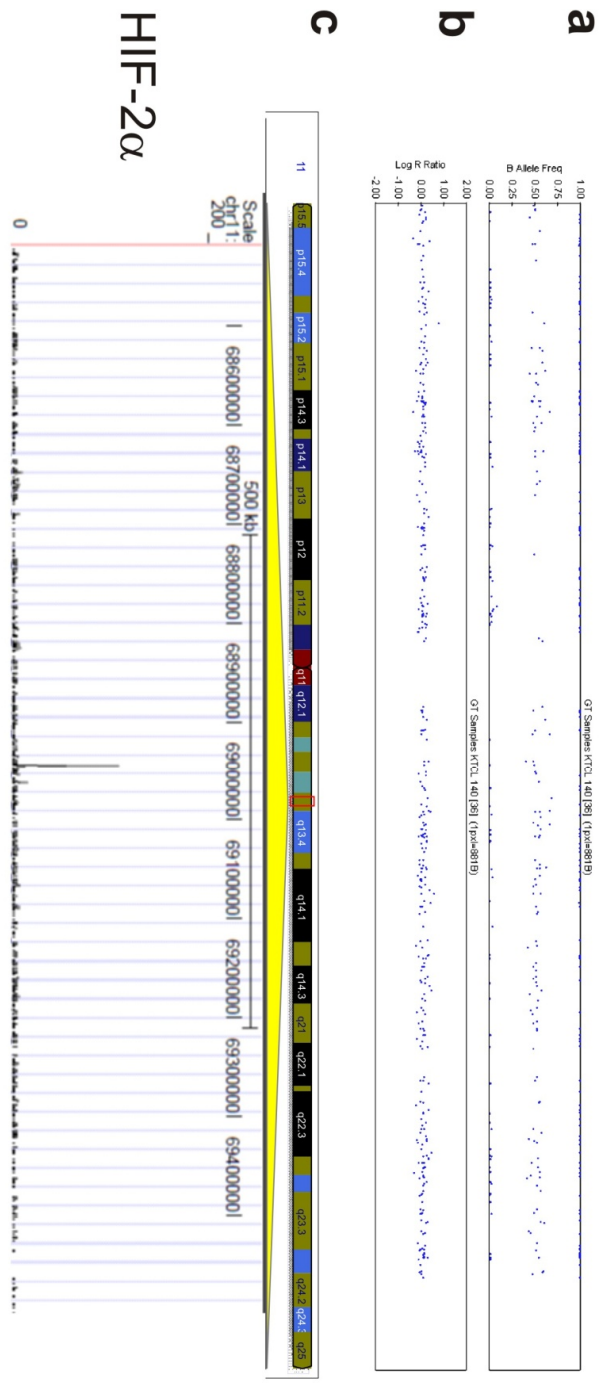


Figure 5.20: OmniExpress SNP array analysis of KTCL140 renal cancer cells. The plots of the region around the 11q13.3 HIF-binding site show (a) retention of heterozygosity indicative of a 50:50 allelic ratio (B allele frequency) and (b) normal copy number relative to the remainder of the genome (Log R ratio) around *CCND1*. (c) HIF-2α ChIP-seq read density corresponding to this region.

Using this cell line, chromatin immunoprecipitations with antibodies directed against HIF-2α and HIF-1β (like 786-O, this cell-line does not express detectable levels of HIF-1α, Figure 5.3) were performed and the resulting material was analysed using Taqman® probes specific for each allele at the rs7948643 locus. Genomic DNA from individuals either heterozygous or homozygous for each allele was included for reference. Input material and chromatin immunoprecipitated with a control antibody showed an allelic ratio comparable to heterozygous DNA (Figure 5.21). Chromatin recovered in both the HIF-2α and HIF-1β immunoprecipitations showed a ratio skewed strongly towards the common/major/risk allele indicating that HIF-2 preferentially binds this allele in vivo and is therefore inhibited from binding the rare/minor/protective allele.

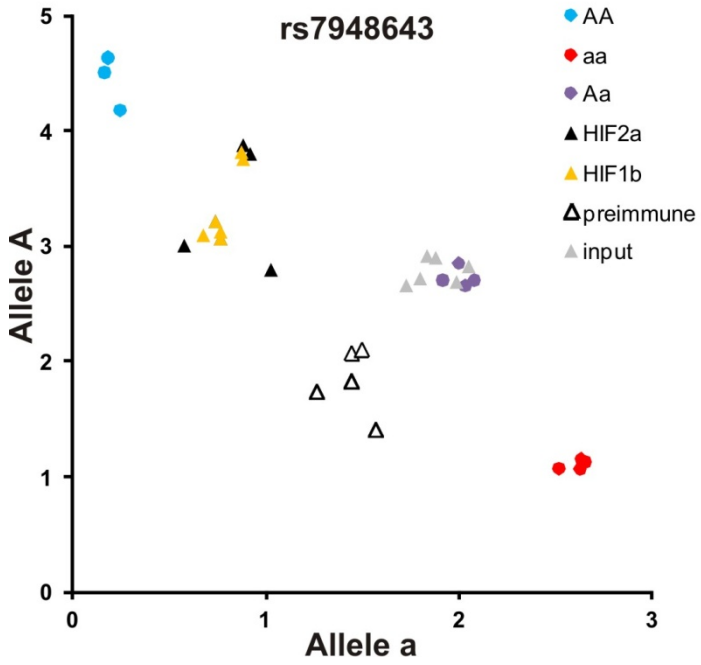


Figure 5.21: The minor (RCC-protective) allele at 11q13.3 disrupts HIF-binding and enhancer activity. Genotype-specific PCR assays at the rs7948643 SNP, performed on immunoprecipitated chromatin from KTCL140 cells. The x-axis shows the intensity of the PCR signal for the minor (protective, a) and the y-axis shows intensity for the major (predisposing, A) allele. Control DNA from individuals homozygous (AA, aa) or heterozygous (Aa) for each allele was included for comparison in every assay. Input material and chromatin immunoprecipitated with preimmune serum (control) showed an allelic ratio comparable to heterozygous DNA. Preferential recovery of allele A by HIF-2α and HIF-1β ChIP (yellow and black) demonstrates preferential binding of both subunits to this allele.

An additional chromatin immunoprecipitation using an antibody directed against RNA pol2 also showed that the common haplotype preferentially interacted with the basal transcriptional machinery at this site indicating that gene expression is predominantly driven by this allele (Figure 5.22a). Thus the rare or protective allele at 11q13.3 disrupts both HIF-binding to the *CCND1* enhancer and HIF-driven gene expression in vivo. Since open chromatin is closely linked to HIF-binding it was then determined whether the two alleles exhibited differing degrees of DNA accessibility. Using a similar approach to that above, material prepared from KTCL140 cells by FAIRE was interrogated using genotype-specific Taqman® probes for the rs7948643 SNP. Open chromatin contained proportionately more of the common allele, indicating that the rare or protective allele is less accessible for transcription factor binding (Figure 5.22b).

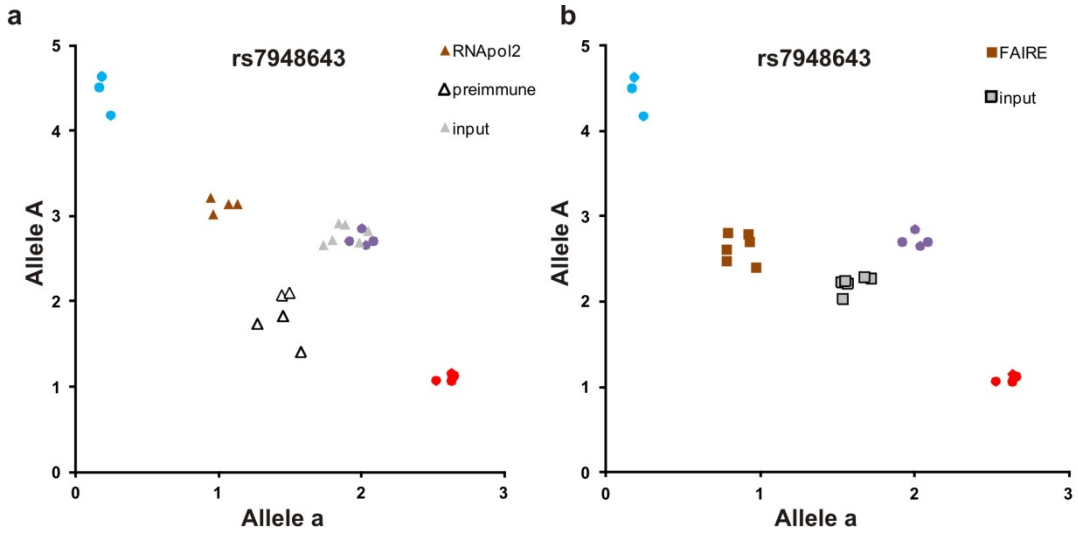


Figure 5.22: (a) Genotype specific PCR assays at the rs7948643 SNP, performed on DNA immunoprecipitated from KTCL140 cells using an antibody directed against RNA pol2 shows greater interaction of allele A with RNA pol2. (b) Genotype specific PCR assays at the rs7948643 SNP, performed on KTCL140 DNA prepared by FAIRE. Preferential recovery of allele A indicates greater DNA accessibility.

5.2.8 Polymorphisms at the enhancer modulate cyclin D1 expression

To examine for allele specific effects on cyclin D1 expression heterozygous coding SNPs (rs7177 and rs678653) within the 3' UTR of the *CCND1* gene were identified. Although

these SNPs are in low linkage disequilibrium with SNPs at the HIF-binding site as determined by the HapMap data and the phase in KTCL140 cells is unknown, mRNA from these cells and DNA recovered by chromatin immunoprecipitation against RNA pol2 show distinct allelic imbalance when compared to genomic DNA, consistent with the hypothesis that allele-specific HIF-binding is driving allele-specific cyclin D1 expression (Figure 5.23). Taken together, this indicates that the protective allele at the 11q13.3 renal cancer associated locus disrupts DNA accessibility, HIF-binding and enhancer activity at a renal cancer-specific promoter-distant neo-enhancer that drives expression of the oncogenic *CCND1* gene specifically in VHL-defective RCC cells.

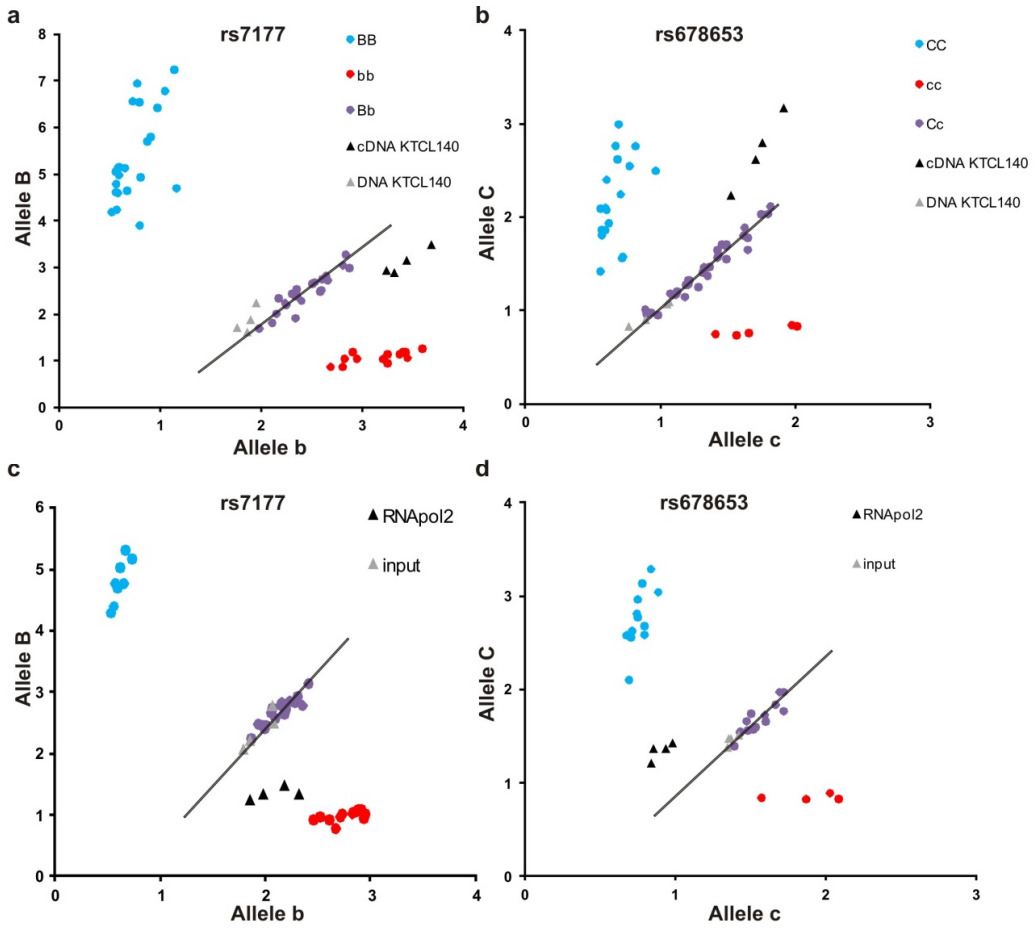


Figure 5.23: Genotype specific PCR assays for rs7177 and rs678653, which lie in the 3'UTR of the *CCND1* gene, using (a and b) cDNA from KTCL140 cells or (c and d) immunoprecipitated DNA using an antibody against RNA pol2 show allele specific expression of *cyclin D1*. Genomic DNA from KTCL140 or control DNA from individuals homozygous (BB, bb, CC, cc) or heterozygous (Bb, Cc) for each allele was included for comparison. Each symbol represents an individual PCR assay. Regression lines were included for comparison.

5.3 Discussion

A recent genome-wide association study has identified two renal cell carcinoma susceptibility loci, one (2p21) that lies within a regulatory region of the gene coding for the transcription factor HIF-2α, which has previously been implicated in ccRCC, and the second of which (11q13.3) lies in an intergenic region of previously unknown function (Purdue, Johansson et al. 2011). Using genome-wide analysis of HIF-binding, together with functional, epigenetic and allele-specific studies, the current work shows that this second locus contains a neo-enhancer that drives expression of the oncogenic gene *CCND1* which is a target of the HIF-2 transcription factor. This provides evidence for a common molecular mechanism that links the two cancer-associated loci identified by the GWAS analysis and ultimately links the VHL tumour suppressor to the cell-cycle regulatory oncoprotein cyclin D1. The activity of this neo-enhancer is inhibited in the rare allele at 11q13.3, which protects against developing ccRCC, with reduction of both DNA-accessibility and HIF-binding. This association between RCC disease and enhancer function, together with the ccRCC-specific pattern of HIF-binding strongly implicates this neo-HIF-binding site in the pathogenesis of renal cancer.

5.3.1 Genetic variants at remote regulatory elements

So far, examples of remote SNPs affecting gene function are sparse. Recently, the SNP rs12913832 which is strongly associated with human pigmentation and which is located in an intron of the non-pigment gene *HERC2* was linked to regulation of the pigment gene *OCA2* which is 21 kb upstream of the SNP (Visser, Kayser et al. 2012). The authors showed that the minor allele of rs12913832 affects chromatin looping, reduces binding of transcription factors to the enhancer and reduces expression of *OCA2*.

Effects on the regulation of an oncogene (similar to the tag SNP rs7105934 examined in this study) have been proposed for SNP rs6983267 on 8q24 which is associated with risk for colorectal and prostatic cancer (Haiman, Patterson et al. 2007; Tomlinson, Webb et al. 2007; Zanke, Greenwood et al. 2007). This SNP is located in a regulatory region about 335 kb telomeric to the *MYC* oncogene and shows long-range interaction with the *MYC* promoter (Pomerantz, Ahmadiyeh et al. 2009; Tuupanen, Turunen et al. 2009). Although experiments showed that the SNP can directly affect TCF4 binding and that the two variants have different activity in reporter analysis a clear effect of genotype on *MYC* gene transcription could not be determined. In the present study the strong association of the VHL-HIF pathway with the regulation of cyclin D1 expression in renal cancer and the allele specific expression in a heterozygous cell line suggest that it is indeed the enhancer on 11q13.3, which confers upregulation of cyclin D1 transcription. A recent report examining the association of tag SNP rs7105934 with renal cancer in a Chinese population confirmed the findings from Purdue et al. (Purdue, Johansson et al. 2011; Cao, Qin et al. 2012). In this study the authors also found an association of lower levels of cyclin D1 mRNA with the presence of the minor allele, although it did not reach statistical significance ($p=0.055$) and the experiments were conducted in normal renal tissue only. Together with the finding that in reporter assays the construct spanning 11q13.3 H + h1 shows higher basal activity this raises the possibility that the enhancer is involved in the regulation of cyclin D1 expression even in the absence of HIF and points to a more complex configuration of transcription factor binding and chromatin looping at this site. Indeed inspection of the extended putative enhancer site 11q13.3 H + h1 for binding of other transcription factors in publicly available data sets on the UCSC browser reveals the presence of other factors (e.g. MAX, FOSL2 or GR) in the (non-renal cancer) cells tested

under normoxic conditions. These factors could be involved in basal and/or HIF-induced regulation of cyclin D1 expression in pVHL-defective renal cancer cells. A more thorough analysis of others factors bound to the enhancer, for example by DNAse footprinting or ChIP analysis in renal cancer cells, could also help to explain the HIF ChIP-seq signal at the minor HIF-binding site 11q13.3 h1, which shows robust enrichment across a number of cell lines, but the sequence does not contain an HRE suggesting that direct binding of HIF to this DNA region is unlikely. A proposed model could be that HIF bound to the major site 11q13.3 H is interacting with other factors bound more downstream thereby creating a loop which after cross linking could be interpreted as HIF-binding signal in the ChIP experiments.

5.3.2 Association of enhancer activity with loss of functional VHL

From the current study it cannot be determined whether VHL loss (and/or high levels of HIF) directly cause this locus to act as a neo-enhancer through some "hard-wired" mechanism, or whether other unconnected (perhaps random) epigenetic events create the neo-enhancer, and VHL loss (and/or high levels of HIF) then provide an advantageous selection pressure that permits outgrowth of these cells. The observation of high cyclin D1 levels which is correlated with VHL loss and HIF activation in early cystic lesions in patients with VHL syndrome (Raval, Lau et al. 2005) indicates that this is an early and common event in the pathogenesis of renal cancer. The data presented in this work suggests that epigenetic mechanisms are strongly involved in regulating accessibility to the 11q13.3 enhancer. This coincides with the finding that modifiers of the epigenetic environment such as the histone modifying enzymes UTX, JARID1C and SETD2 or the chromatin remodelling PBRM1 are often mutated in ccRCC (van Haaften, Dalgliesh et al. 2009; Dalgliesh, Furge et al. 2010; Varela, Tarpey et al. 2011; Guo, Gui et al. 2012; Larkin,

Goh et al. 2012) and underscores the importance of epigenetic modifications in the pathogenesis of RCC. The ongoing operation of this enhancer despite the stable re-introduction of VHL (with consequent HIF-suppression) over many generations of cell division, indicates that VHL loss (and/or high levels of HIF) are not required to maintain this enhancer. Thus if a direct action of VHL or HIF did lead to generation of the neo-enhancer it would only be as an initiator. Since high levels of cyclin D1 are seen in the early cysts of VHL patients such an initiating event must happen very early in tumour progression (Raval, Lau et al. 2005).

Whilst HIF-binding to this enhancer is not observed in non-malignant renal tubular cell lines or in cancer cell lines from other tissues, a physiological role for this enhancer in other tissues or during development cannot be ruled out. However, it should be noted that the HRE at the centre of the main HIF-binding peak is not conserved in non-primate mammals such as mice, a feature which may contribute to the resistance of VHL knockout mice to developing renal tumours.

5.3.3 Implications for further studies

In this work, analysis of the effects of HIF-binding to the 11q13.3 enhancer was focused on the regulation of the oncogene *CCND1* which has been shown to be strongly regulated by HIF in the context of renal cancer. However, effects of this enhancer on other (maybe yet unknown) genes cannot be ruled out. Therefore, a study of genome-wide interactions of this enhancer or other relevant HIF-binding sites using techniques such as 4C (Chromosome Conformation Capture-on-Chip) or ChIA-PET (Chromatin Interaction Analysis by Paired-End Tag sequencing) would be of great interest and has been successfully applied to the estrogen receptor pathway (Fullwood, Liu et al. 2009). This could further define the network involved in renal cancer development.

The co-localization of RCC-specific HIF DNA-binding with an RCC susceptibility locus raises the possibility that additional RCC-specific HIF-binding sites may also be important in the pathogenesis of renal cancer. These may contribute to cancer-susceptibility, but be below the limit of detection of the former study, or they may pertain to other aspects of cancer biology such as the response to treatment, prognosis or metastasis that were not examined in the published GWAS. Furthermore, since such sites may lie at great distances from coding regions, they may not be detected by exomic approaches at identifying cancer associated loci such as those being undertaken by the **C**atalogue **O**f **S**omatic **M**utations **I**n **C**ancer (COSMIC). A ChIP-seq approach across a panel of ccRCC and non-cancerous cell lines would provide a rapid way of detecting such sites. Conversely, this work provides a paradigm for the analysis of other disease associated intergenic loci of unknown function, through cross-referencing with the rapidly expanding ChIP-seq databases of genome-wide transcription factor binding, DNA accessibility and chromatin modification.

6 CHAPTER 6: Final Discussion

The lack of oxygen in cells leads to an altered metabolic environment challenging the survival of cells. Therefore, measures within the cells to adapt to these unfavourable conditions are essential for survival. On the level of gene transcription most of these responses are mediated via HIF which directly controls the upregulation of hundreds of genes involved in adjusting the energy balance to lower oxygen concentrations.

In this work, I have characterised the genomic environment of HIF DNA-binding sites to better understand basic mechanisms of the hypoxic response. In each of the two cell lines examined and consistent with the work of others, HIF binds to about 500 sites in the whole genome when applying strict criteria. Interestingly, a large number of these sites are distant from annotated promoters in non-coding regions of the genome. These sites mainly reflect the existence of a cell type specific repertoire of enhancers modulating the expression of genes important for the hypoxic response. Some of these remote sites also control the expression of as yet unidentified transcripts with unknown functions. The results from this study are in line with recent findings from the ENCODE consortium which aimed to map regulatory regions in the whole genome (Dunham, Kundaje et al. 2012). In fact over 80% of the genome exhibit characteristics of regulatory elements, defined as RNA- or chromatin-associated events in at least one of the cell types examined. As opposed to the 2.94% of the genome covering exons of protein-coding genes or just 1.22% of DNA covering protein-coding exons this vast part of the genome offers vast possibilities for gene regulation. Therefore, the finding that HIF binds to remote regulatory elements and that differences between cell lines are more pronounced at these sites is a common feature of gene regulation. The fact that in addition to HIF-binding events, chromatin changes occur in hypoxia adds another level of complexity.

It is plausible that regulatory regions are in contact with more than one promoter on more than one chromosome. This would result in a combinatorial network which is dependent on the cellular background, the three dimensional folding of the chromatin and the fourth dimension – time, since transcription factor binding is not necessarily permanent and strongly depends on the functional presence of transcription factors (e.g. HIF is only present in hypoxia). In respect to the HIF system, most of the regulatory elements which are able to bind HIF are already in an open chromatin conformation in normoxic conditions, suggesting that beyond a small number of core hypoxia response regulatory elements each cell type possesses a specific program to respond to a hypoxic insult with a different portfolio of genes regulated. However, the mechanisms involved in establishing this status remain obscure. Mapping HIF-binding sites in more cell types, generating a HIF-interactome within the nucleus and integrating additional hypoxic driven events clearly provides challenging but rewarding tasks for future experiments.

A multitude of polymorphisms are identified in GWAS analysis in the context of a variety of diseases. Most of these polymorphisms lie in intergenic regions with no known relation to genes involved in the respective disease. Therefore, it is likely that most SNPs are affecting remote regulatory regions which are in turn involved in the regulation of genes far distant from the identified locus. This is not necessary the next nearest gene as exemplified in this work since the *CCND1* gene is not the closest gene to the 11q13.3 enhancer. GWAS analyses provide the basis for further functional experiments related to HIF effects in cancer biology. More SNPs than the ones reported are associated with the development of renal cancer and need to be analysed in respect to local HIF-binding. It is also plausible that additional as yet unidentified genotypes are associated with clinical features of renal cancer such as metastasis, time of survival or response to therapy.

Correlating genotypes e.g. at the 11.q13.3 *CCND1* enhancer with the clinical course would give more insights to the role of SNPs and HIF in renal cancer biology. Furthermore, the general role of HIF in cancer biology raises the questions whether loci identified in GWAS studies conducted in populations with other diseases need to be examined for HIF-binding events.

The present study provides a methodological basis for analysing genome-wide HIF-binding data, first insights into the DNA and chromatin environment at HIF-binding sites, a thorough analysis of hypoxic gene regulation and first evidence that remote HIF-binding to the 11.q13.3 enhancer is crucially involved in the development of clear cell renal cancer.

7 CHAPTER 7: Material and Methods

7.1 Cell lines

786-O, MCF-7, HK-2 and Hela cell lines were purchased from American Type Culture Collection (ATCC, via LGC, U.K.). HKC-8 cells were a kind gift from L. Racusen, Baltimore, USA and were received from K.-U. Eckardt, Erlangen, Germany. All other cell lines used for this study were thawed from the cryo-stock cell collection of the laboratory.

The karyotypes of 786-O and MCF-7 cell lines (which were used for FISH, 3C assays, FAIRE assays and chromatin immunoprecipitations) were examined by Dr. Emanuela Volpi and Dr. Mohammed Yusuf (Molecular Cytogenetics and Microscopy Core, WTCHG, University of Oxford, UK). Furthermore, copy numbers of the 11q13.3 locus in these cell lines were determined by Dr. Mohammed Yusuf using fluorescent labelled BAC clones (probe A = RP11 266K14 and probe B = RP11 378K8) and fluorescence in situ hybridisation.

786-O is a near-tetraploid cell line, with a substantially rearranged karyotype although chromosome 11 appeared structurally unaffected. Four copies of chromosome 11 were observed in the majority of cells analysed (54 in total, 22 informative metaphases with probe A and 32 informative metaphases with probe B). Each chromosome 11 homolog showed one signal with both probe A and probe B (Figure M1a).

MCF-7 is a near-triploid, highly rearranged cell line. Regarding chromosome 11 two seemingly unaltered copies plus either 3 or 4 derivative chromosomes with some chromosome 11 material were detected (55 cells in total analysed, 18 informative metaphases with probe A and 37 informative metaphases with probe B). The two unaffected homologues plus three derivatives show one signal with probe A and probe B (Figure M1b).

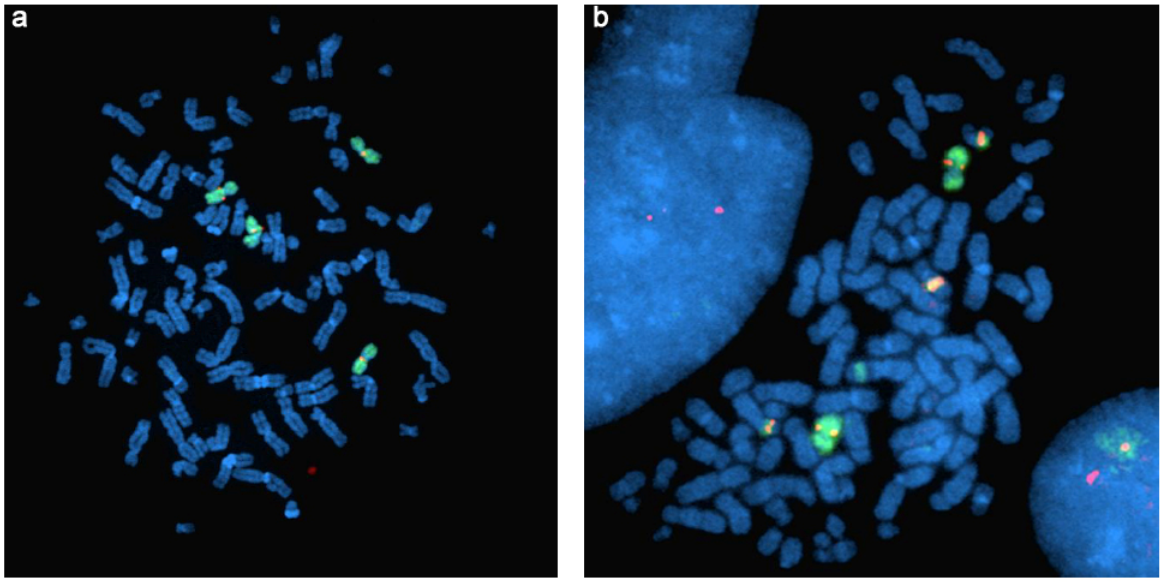


Figure M1: (a) 786-O cell metaphase labelled with marker for chromosome 11 (green) and probe A (red) covering parts of the 11q13.3 locus. All four copies of chromosome 11 show signals for probe A. (b) MCF-7 cell metaphase labelled with marker for chromosome 11 (green) and probe A (red). Two seemingly structurally unaffected copies of chromosome 11 show two signals of probe A. From four derivatives of chromosome 11 three are labelled with probe A.

7.2 Cell Culture

HKC-8 cells were cultured in Dulbecco's modified Eagle's medium/Ham's F-12 supplemented with 10% fetal calf serum, 2 mM L-glutamine, 100 U/ml penicillin and 100 µg/ml streptomycin, 5 µg/ml insulin, 5 µg/ml transferrin, and 5 ng/ml selenium (Sigma, U.K.). All other cell lines were grown in Dulbecco's modified Eagle's Medium, 100 U/ml penicillin, 100 µg/ml streptomycin and 10% fetal bovine serum (Sigma, U.K.) at 37 °C in a cell culture incubator. All experiments with reduced oxygen concentrations were conducted in a Ruskin Invivo2 400 work station equipped with a Gas Mixer Q (Ruskinn, U.K.).

7.3 DNA-techniques

7.3.1 General DNA techniques

7.3.1.1 Isolation of genomic DNA

1. 10cm tissue culture dishes of sub confluent cells were washed once with PBS, and the cells were scraped into 1 ml of PBS and pelleted in a 1.5 ml tube at 4 °C at 2500 rpm. In general, centrifugations of 1.5 ml tubes were performed in a Hettich micro centrifuge (MIKRO 200R, Rotor 2424-A, Radius 85mm).
2. PBS was removed and cells were lysed in 0.5 ml cell lysis buffer for DNA extraction at 55 °C over night.
3. After transferring the samples into a fresh 1.5 ml tube 0.5 ml of phenol chloroform was added, samples were vortexed for approximately 10 seconds and centrifuged at 14000 rpm for 10 min at 4 °C.
4. The upper phase was transferred into a fresh 1.5 ml tube and DNA was precipitated by adding 50 µl of 3M NaOAc and 1 ml of 100% Ethanol.
5. Samples were frozen at -80 °C over night.
6. After centrifugation at 14000 rpm for 45 min at 4 °C the DNA pellet was washed once with 1 ml of 70% Ethanol, centrifuged again for 30 min.
7. The supernatant was removed and the DNA pellet air dried for approximately 10 min.
8. DNA was resuspended in 0.05-0.2 ml DNAse free H_2O.

7.3.1.2 Quantification of DNA concentration

Double stranded DNA concentration was determined by measuring the absorbance of the DNA solution at 260 nm using a NanoDrop® ND-1000 spectrophotometer (Thermo Scientific, U.K.). 1.5 µl of DNA was measured against 1.5 µl water control and DNA was considered pure at absorbance ratios 260/280 higher than 1.8.

7.3.1.3 Polymerase chain reaction (PCR) amplification of DNA

PCR was performed in 50 µl reactions using a PTC-200 Peltier thermal cycler (MJ Research, U.K.).

PCR reactions using the Bioline PCR kit (Bioline, U.K.) were as follows:

Name	**Stock**	**Volume**
MgCl	50 mM	2.5 µl
PCR-buffer	10X	5 µl
Primer fw	10 µM	1 µl
Primer rev	10 µM	1 µl
dNTPs	25 µM	4 µl
Template DNA	30 ng/µl	2 µl
DNA Polymerase	5 U/µl	0.5 µl
H_2O		34 µl

Typical PCR cycle conditions were as following:

Step	**Temp in °C**	**Time**	**Repeats**
1	95	3 min	1
2	95	30 sec	
	58	30 sec	
	72	1 min	30
3	72	5 min	1
4	4		

7.3.1.4 Restriction digest of DNA

1 µg of plasmid DNA or 15 µl of a PCR reaction were digested for 2 h at 37 °C using 10-20 U of the appropriate restriction enzyme (NEB, U.K.) and the recommended buffer. Conditions were adjusted if a double digest was performed.

Alkaline phosphatase treatment to remove 5' phosphate groups was performed by adding 10 U of calf intestinal alkaline phosphatase (NEB, U.K.) for 30 min at 37 °C. Restriction digest efficiency was verified by running the digested sample on an agarose gel. Bands were then cut out from the gel and purified as below.

7.3.1.5 Gel electrophoresis

1-2% agarose gels were prepared by melting agarose in TRIS/Borate/EDTA (TBE) buffer. After cooling to approximately 50 °C, ethidium bromide (4 µl / 100 ml solution) was added and the gel was cast. Samples were mixed with (6x) loading buffer and electrophoresis was performed using a Biorad gel electrophoresis chamber (Biorad, U.K.) at 100V for 20-30 min. Size of the band were compared to marker DNA bands (100 bp marker or 1 kb marker, NEB, U.K.) and were documented using the AutoChemi Bioimaging System from UVP (U.K.). Bands of interest were excised under low intensity hand-held ultraviolet light.

7.3.1.6 Purification of amplified or digested DNA

PCR products (either from gels or directly from the PCR reaction), sheared genomic DNA or digested vectors were purified using the GFX® PCR DNA and gel band purification kit (GE Healthcare, U.K.) according to manufacturer's instructions.

7.3.1.7 Ligation of DNA fragments

Ligation of digested vector constructs and digested inserts was performed using T4 DNA ligase (NEB, U.K.). A threefold molar excess of insert DNA over vector was used. The reaction was prepared as outlined below in 1.5ml tubes and ligations were carried out for either 2h at room temperature or over night at 16 °C.

Ligation	**Source**	**Stock**	
Buffer	NEB	10X	1 µl
Ligase	NEB	10 µM	1 µl
H_2O			6 µl
Vector DNA			1 µl
Insert DNA			1 µl

7.3.1.8 Generation of vector constructs for transfection

Sequences to be studied were inserted into pGL3 basic and pGL3 promoter constructs (Promega, U.K., Figure M2) using cloning sites in their polylinker regions as specified.

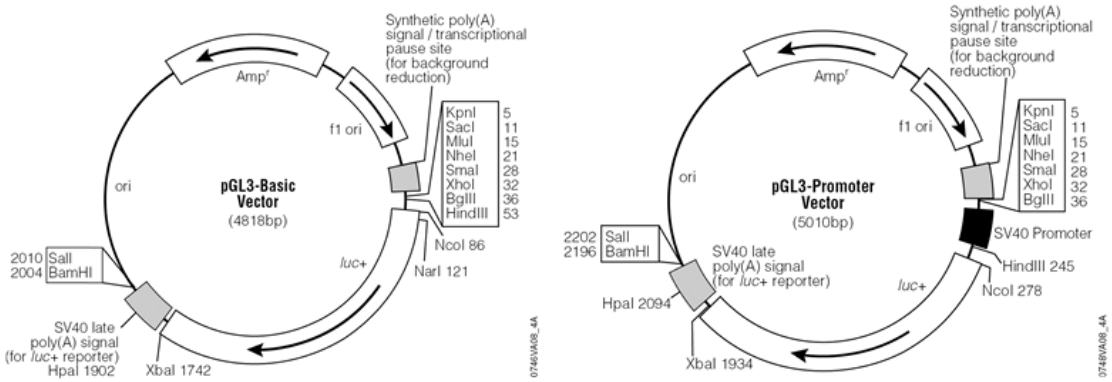


Figure M2: Vector maps of the pGL3 basic and the pGL3 promoter vectors. Downloaded from: http://www.promega.com

7.3.1.8.1 Bacterial strains, transformation and isolation of constructs

For transformation steps XL-10-Gold ultra-competent cells (Stratagene, U.K.) were used according to the manufacturer's instructions.

1. Cells were thawed on ice for 30 min.
2. 20 µl of cell suspension was pipetted into a 1.5 ml tube and 2 µl of ligation sample was added.
3. The samples were kept on ice for 10 min before heat shock treatment at 42 °C for 45 sec in a water bath.
4. Samples were kept on ice for 2 min and then 500 µl of SOC medium (Invitrogen, U.K.) was added and samples were incubated at 37 °C rotating at 200 rpm for 45 min.
5. Cells were centrifuged 5 min at 2000 rpm, the supernatant was removed to about 50 µl and cells were resuspended and plated onto a LB agarose plate containing ampicillin (0.1 mg/ml) for overnight growth.
6. Control ligations containing only the vector construct but not the inserts were processed in parallel to estimate background growth.
7. A selection of colonies were picked from the LB agarose plate and transferred into 5 ml of LB medium containing ampicillin (0.1 mg/ml) in a 30 ml vial and incubated at 37 °C shaking with 200 rpm for 16 h.
8. For large scale amplification the 5 ml bacterial cultures were diluted with 200 ml of LB medium containing ampicillin (0.1 mg/ml) and incubated for further 16 h.
9. Cells were collected by centrifugation at 3750 rpm at 4 °C for 10 min in a Beckman Coulter centrifuge (Beckman Coulter, U.K.). This centrifuge was used for all 15 ml and 50 ml tubes (Beckman Coulter, Allegro X-12R, Rotor SX-4750, Radius 207.8

mm). Plasmids were isolated using plasmid purification kits from Qiagen following manufacturer's instructions (Qiagen, U.K.). For small scale purifications the QIAprep spin miniprep system was used, for larger scale purifications the Hi speed plasmid maxi kit was used (Qiagen, U.K.).

7.3.1.8.2 DNA sequencing

DNA sequences of vector constructs were verified using a commercial sequencing facility in Oxford (Source Bioscience, U.K.).

7.3.1.8.3 Cloning of specific constructs

Inserts amplified from genomic DNA from 786-O cells spanning 227 bp (HG18 Chr11: 68943675-68943901) of the 11q13.3 H major peak or 1453 bp (HG18 Chr11: 68943675-68945127) including the 11q13.3 H major and the 11q13.3 h1 minor peak into the BGL2 restriction site of the pGL3 promoter construct. For the 11q13.3 h2 peak a 556 bp insert (HG18 Chr11: 68961017-68961572) was cloned into the NHE1/KSP1 restriction site of the pGL3 promoter vector. Mutations of HREs (RCGTG into RCATA) were performed using site directed mutagenesis (see below).

Sequences covering the HIF-binding sites for *ARRDC3* (276 bp) and *ERRFI1* (the main peak, 278 bp) were cloned into pGL3 promoter vector using the BglII restriction enzyme. A 682 bp sequence spanning the HIF-binding site close to SLC2A3 ("Nici") was cloned into the pGL3 basic vector using the NHE1/KSP1 restriction sites.

Restriction enzyme digests and ligation reactions were carried out as described above.

7.3.1.8.4 Site directed mutagenesis (SDMG)

Site directed mutagenesis (SDMG) was performed using vector constructs as templates and oligonucleotides with the desired base exchanges for PCR amplification.

The SDMG PCR reaction was prepared as below:

Name	**Stock**	**Volume**
PFU buffer	10X	5 µl
dNTP Mix	8 mM	6.25 µl
DMSO	100%	5 µl
Primer fw	10 µM	2 µl
Primer rev	10 µM	2 µl
Template Plasmid	100 ng/µl	0.5 µl
PFU polymerase	2.5 U/l	1 µl
H_2O		28.25 µl

PCR cycle conditions in a PTC-200 Peltier thermal cycler (MJ Research, U.K.) were as following:

Step	**Temp in °C**	**Time**	**Repeats**
1	95	30 sec	1
2	95	30 sec	
	56	1 min	
	68	7 min	16
3	4		

SDMG PCR was followed by adding 1 µl DPNI restriction enzyme (10 u/µl, Fermentas, U.K.) to digest the methylated template plasmid and by incubating the sample for 30 min at 37 °C. Products were run on an agarose gel and mutated plasmids were cut out from the gel. Finally, the mutated insert was excised from the plasmid using restriction enzyme digest, gel purified and religated back into the original vector plasmid (that had not been subjected to SDMG PCR) to avoid the possibility of random PCR generated mutations in the vector back bone.

7.3.1.8.5 Transfection of plasmid DNA

Transfections of plasmids were performed in 786-O cells or U2OS cells using Fugene6® reagents (Promega, U.K.). Cells were harvested 48 h after transfection and luciferase activity was measured using a Luciferase reporter system (Promega, U.K.).

1. Cells were plated in 6 cm dishes on the day before starting transfection and reached 30-50% confluency on the day of transfection.
2. Typically 3 µg of plasmid vector DNA and 1 µg of transfection control plasmid (beta-galactosidase, renilla luciferase) were mixed in 0.5 ml Optimem medium.
3. In a second tube 12 µl of Fugene6® was mixed with 0.5 ml Optimem.
4. Both solutions were vortexed and kept for 5 min at room temperature. The plasmid mix was then transferred into the transfection reagents mix, vortexed and kept for 15 min at room temperature.
5. 0.5 ml of the transfection mix was directly added to 2 ml of medium with the cells.
6. 786-O cells were incubated under normoxic conditions and harvested 48 h after transfection.

7. U2OS cells were split 24 h after transfection into 12-well dishes and aliquots of the cells exposed to either normoxia or hypoxia (0.5 % O_2) for 16 h before cell harvest.
8. U2OS cells, cotransfected with the pGL3 reporter and β-galactosidase reporter, were cultured in normoxia or 0.5% hypoxia for 16 h. Luciferase activity was normalized to β-galactosidase activity.

All constructs were sequence verified.

7.3.1.8.6 Reporter assays

Reporter assays were performed using the Luciferase Reporter assay (Promega, U.K.) or the Dual-Luciferase Reporter assay (Promega, U.K.) exactly following manufacturer's instructions.

1. 12-well dishes of cells, transfected as above, were washed with PBS, lysed in 100 µl Passive lysis buffer (Promega, U.K.), frozen and stored for at least 2 h at -80 °C.
2. Cells were scraped and the lysed cells were transferred to a 1.5 ml tube.
3. Samples were centrifuged at 14000 rpm for 5 min and 20 µl of the supernatant was pipetted into one well of a U-bottomed white 96-well plate.
4. Firefly luciferase activity was determined using a Luminoskan Ascent luminometer (Labsystems, U.K.) which added 50 µl of Luciferase Assay Substrate (Promega, U.K.) and measured activity.
5. In Dual-Luciferase experiments renilla luciferase activity was measured in parallel for normalisation using the same settings but the provided Stop&Glo® substrate (Promega, U.K.). Firefly luciferase activity was corrected for transfection efficiency by dividing the reading by the respective renilla luciferase activity reading.

6. In experiments were beta-galactosidase was used for determining transfection efficiency beta-galactosidase activity was measured in 20 µl cell extract by adding 80 µl Z-buffer to the lysate in each well of a clear U-bottomed 96-well plate.
7. This was followed by adding 20 µl of o-nitro-phenyl-beta-D-galactopyranoside (4mg/ml) and incubation of the plate at 37 °C with shaking at 200 rpm. On the appearance of a yellow colour, reactions were stopped by adding 50 µl Na_2CO_3 (1M).
8. Absorbance was measured at 405 nm using a Vmax kinetic microplate reader (Molecular Devices, U.K.). Luciferase activity was corrected for transfection efficiency by dividing the luciferase activity reading by the respective beta-galactosidase reading. In general single reporter assays were performed on triplicate biological samples and mean values and standard deviation were calculated.

7.4 Protein Techniques

7.4.1 Protein lysates

1. Cells were cultured in either 10 cm dishes or 6-well plates and harvested at sub-confluency for protein isolation.
2. Medium was removed and cells washed once with ice-cold PBS.
3. Cells were scrapped in 5 ml (10 cm dish) or 1 ml (6-well) ice-cold PBS, transferred to a 15 ml or 1.5 ml tube, respectively, and collected by centrifugation at 2000 rpm for 5 min at 4 °C.
4. Supernatant was removed and cells lysed in 100-200 µl Urea/SDS buffer.

5. Protein content was measured using the DC protein assay (Biorad, U.K.) following manufacturer's instructions performed in triplicates per assay. Protein concentrations were normalised between the different conditions by dilution with Urea/SDS.

7.4.2 Immunoblotting

SDS-PAGE (sodium dodecyl sulphate polyacrylamide gel electrophoresis) was used to separate proteins of interest. Resolving acryl amide gels were prepared using SDS resolving gel buffer with 30% acryl amide:bisacryl amide mix (37.5:1, BioRad, U.K.) to percentages of 8-15% depending upon the molecular weight of the protein to be analysed. Stacking gels were prepared with SDS stacking buffer and addition of 30% acryl amide:bisacryl amide mix to a final concentration of 3.9%. Polymerisation of the gels was achieved by adding ammonium persulfate (0.1% w/v final, Sigma, U.K.) and N,N,N',N'-tetramethylethylendiamine (0.1% w/v, final).

1. SDS-PAGE was performed in a Mini-Protean apparatus (BioRad, U.K.). Loading buffer to 1X was added to the samples and samples were heated to 95 °C for 5 min.
2. Typically 10-20 µl of sample was loaded per well and a protein standard - Page ruler plus (Fermentas, U.K.), was run on the same gel.
3. Electrophoresis was carried out in SDS-PAGE running buffer at a constant voltage of 120V for 1 h - 1 h 30 min.
4. Proteins were then transferred onto polyvinylidene fluoride (PVDF) membrane (Millipore, U.K.) using the Mini-Transblot apparatus (BioRad, U.K.). The membrane was pre-activated for 1 min in 100% methanol, washed and pre-equilibrated for

10 min in 1X transfer buffer. Transfer was carried out at constant voltage of 110V for 1 h with ice-cooling.

5. Thereafter membranes were transferred into 5% non-fat milk blocking solution for at least 30 min.
6. Membranes were incubated with the specified primary antibody (diluted in 5% non-fat milk blocking solution) over night at 4 ^{o}C.
7. Membranes were washed three times for 10 min each in 5% non-fat milk blocking solution before incubation with HRP-coupled secondary antibodies (anti-mouse IgG or anti-rabbit IgG, 1:2000 in 5% non-fat milk, Dako, U.K.) for 45 min.
8. Membranes were washed once with 5% milk followed by two washes with 1X PBS-T for 10 min.
9. Immunodetection was carried out by visualising the HRP activity using the chemiluminescence reagents SuperSignal West Pico, Dura or Femto (Pierce Biotechnology, U.K.) as applicable.
10. X-OMAT LS films (Kodak, U.K.) were exposed to the membranes to detect chemiluminescence.
11. For quantitation of the signal the BioSpectrum AC imaging system with LabWorks software (UV Products, U.K.) was used.

7.5 RNA techniques

7.5.1 RNA isolation from cultured cells

RNA from cells was isolated using phenol-chloroform based extraction methods (Trizol® reagent, Sigma, U.K.) according to the manufacturer's protocol.

1. Cells were grown in 6-well dishes to about 90% confluency. Cells were washed with ice cold PBS and kept on ice.
2. 0.5 ml of Trizol® reagent was added and the cell suspension pipetted up and down five times before transferring the suspension to 0.5 ml tubes. Samples were kept at room temperature for 5 min.
3. 100 µl of chloroform were added and the sample shook by hand for 15 sec.
4. Samples were centrifuged at 14000 rpm at 4 °C for 20 min and the supernatant transferred to a new 0.5 ml tube.
5. 250 µl of isopropanol were added and the sample kept at -80 °C overnight.
6. Samples were centrifuged at 14000 rpm at 4 °C for 45 min, supernatant was removed and the pellet washed with 0.5 ml of 75% ethanol.
7. Samples were centrifuged at 14000 rpm at 4 °C for 30 min, supernatant was removed and the pellet air dried for approximately 10 min.
8. RNA was resuspended in 25 µl of RNAse free water and kept at -80 °C.

7.5.2 Control of RNA quality

RNA was dissolved in RNAse free water and concentration and purity was determined using the NanoDrop analyser (Thermo scientific, U.K.). Absorbance ratios 260/280 of 1.9 or higher RNA were considered adequate to use the RNA for further analysis.

7.5.3 Generation of complementary DNA

cDNA was generated using the High capacity cDNA kit (Applied Biosystems, U.K.) according to the manufacturer's instructions. 500-1000 ng of RNA were reverse transcribed in a PTC-200 Peltier thermal cycler (MJ Research, U.K.). For subsequent experiments cDNA reactions were diluted 1 in 5 with DNAse free water.

cDNA reaction mix:

Name	**Stock**	**Volume**
10x RT Buffer	10X	2 µl
25x dNTP Mix	100 mM	0.8 µl
10x RT Random Primers	10x	2 µl
MultiScribe® Reverse Transcriptase		1 µl
RNAse Inhibitor		1 µl
RNA template	500 ng/µl	1 µl
Nuclease free H_2O		12.2 µl

cDNA reaction conditions:

Step	**Temp in °C**	**Time**
1	25	10 min
2	37	120 min
3	85	5 min

7.5.4 Expression and enrichment analysis using Taqman® or Sybr green® assays

Expression analysis was performed using Taqman® assays (Applied Biosystems, U.K.) or the SYBR green® (Applied Biosystems, U.K.) method. Primer design for the SYBR green® method was performed using the publically available primer3 web tool (http://frodo.wi.mit.edu/). Since chromatin sonication resulted in DNA fragments of

between 300 and 500 bp qPCR product size was restricted to between 60 bp and 150 bp. For assays of cDNA, primers spanning at least two exons and a large intron (>5000 bp) were designed in order to avoid amplification of potentially contaminating genomic DNA. PCR primer specificity was checked routinely using the *in silico* PCR tool at the UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgPcr?command=start). PCR was performed using the StepOnePlus® Real-time PCR system (Applied Biosystems, U.K.) using the following settings.

7.5.5 Settings for qPCR experiments

Taqman® method

Step	Temp in °C	Time	Repeats
1	95	20 sec	
2	95	1 sec	
	60	20 sec	40

SYBR green® method

Step	Temp in °C	Time	Repeats
1	95	20 sec	
2	95	3 sec	
	60	30 sec	40
3	95	15 sec	
4	60	1 min	
5 Melting curve	Increase of +0.3 per step to 95		

qPCR reaction mixes

Taqman® assay	**source**	**Stock**	**Volume**
PCR buffer	AB	2x	5 µl
Primer and probe mix	AB	20x	0.5 µl
Control primers and probe mix	AB	20x	0.5 µl
H_2O			2 µl
cDNA/DNA			2 µl

SYBR green® assay	**Source**		
Fast SYBR green® PCR buffer	AB	2X	5 µl
Primer mix	Sigma	10 µM	0.5 µl
H_2O			2.5 µl
cDNA/DNA			2 µl

7.5.6 Data analysis

Expression or enrichment levels were calculated as fold regulation/enrichment compared to the specified control using the ΔΔct method. Threshold cycles (ct) of the transcript of interest were compared to ct values of a housekeeping transcript (HPRT, Cyclophyllin A) and then normalised to the external reference sample (e.g. normoxia, untreated, control). Ct values higher than 35 cycles were excluded from calculation. For SYBR green® assays a melting curve analysis was performed to verify single products in the PCR reaction.

7.5.7 Microarray analysis

Quality of RNA for microarray analysis was checked using the Bioanalyzer 2100 (Agilent, U.K.) according to the manufacturer's instructions. RNA samples with RNA integrity numbers (as defined by the manufacturer) of over 9 were considered of sufficient quality for microarray analysis. Amplification and hybridization of cDNA to the Human WG6v3 expression bead chip (Illumina, U.K.) were performed by Dr. Dilair Baban from the microarray facility at the Wellcome Trust Centre for Human Genetics, Oxford, U.K. Each condition was performed in biological triplicate and signal intensities were quantile normalised and filtered for present/absent calls by the Illumina software. The regulation of each gene was then expressed as a ratio of the mean expression level between each condition.

7.5.8 Transfection of silencing RNA (siRNA)

siRNA Name	**Source**	**Sense**	**antisense**
HIF-1	Ambion	CUGAUGACCAGCAACUUGAtt	UCAAGUUGCUGGUCAUCAGtt
HIF-2	Ambion	CAGCAUCUUUGAUAGCAGUtt	ACUGCUAUCAAAGAUGCUGtt
dHIF	Ambion	CCUACAUCCCGAUCGAUGAtt	UCAUCGAUCGGGAUGUAGGtt
siRNA Target	**Source**	**NAME**	**Cat.- No.**
ARNT	Dharmacon	On Target plus Smart pool ARNT	L-007207-00-0005
Non-targeting	Dharmacon	On Target plus non-targeting pool	D-001810-10-20

siRNAs targeting human HIF-α or drosophila HIF (dHIF, control) were transfected into cells using oligofectamine reagent (Invitrogen, U.K.) according to the manufacturer's instructions.

1. Cells were plated into 6-well plates the day before starting transfections and reached 30-50% confluency on the day of the first transfection. Typically two transfections of siRNA were carried out 24 h apart.
2. For one well 0.3 ml of Optimem medium (Invitrogen, U.K.) were mixed with 3 µl oligofectamine in a 1.5 ml tube. In a separate tube 1.2 ml Optimem medium were mixed with 1.5 µl of siRNA (40 µM) in order to reach a final concentration of 40 nM in a total volume of 1.5 ml. Both tubes were vortexed briefly and kept at room temperature for 5 min.
3. The oligofectamine mix was transferred into the siRNA mix, the tube was vortexed briefly and kept 15 min at room temperature.
4. The cells were washed three times with PBS in order to remove traces of serum, which could inhibit transfection and the 1.5 ml mix of transfection reagents was pipetted onto the cells.
5. Cells were incubated for 4 h at 37 °C before adding 0.5 ml 40% fetal bovine serum in Optimem medium to the cells.
6. The transfection was repeated after 24 h and cells were harvested 48 h after the first transfection.

For VHL-defective cell lines the experiments were carried out under normoxic conditions. For VHL competent cells parallel experiments were conducted and cells were exposed either to normoxia or hypoxia for 16 h prior to cell harvest. Cells were then processed for isolation of RNA or proteins as applicable.

7.6 Specific DNA techniques

7.6.1 Preparation of BAC, FOSMID and Image clones

Commercially sourced clones used in this study were:

Clone ID		Source	Genomic region covered
G248P85537B2	Fosmid	BACPAC	hg18 chr11: 69147108-69190642
G248P83053H10	Fosmid	BACPAC	hg18 chr11: 68921622-68961904
G248P85591A1	Fosmid	BACPAC	hg18 chr11: 68697196-68733066
G248P89319G2	Fosmid	BACPAC	hg18 chr11: 69380087-69418798
RP11 266K14	BAC	Source Bioscience	hg18 chr11: 69145503-69298443
RP11 378K8	BAC	Source Bioscience	hg18 chr11: 68913046-69094413
NIH_MGC_71cDNA clone IMAGE:5549618	Image clone	Source Bioscience	Length 1060 bp

Bacterial clones were cultured in selection medium according to the instructions provided by the source company. For BAC clones BAC DNA was isolated from 250 ml cultures using the Qiagen Large-construct kit following manufacturer's instructions (Qiagen, U.K.).

Fosmid DNA was isolated as follows. 5 ml of LB medium supplemented with the appropriate antibiotic (ampicillin 50 µg/ml or chloramphenicol 170 µg/ml) were inoculated with the respective bacterial strain containing the fosmid and cultured at 200 rpm at 37 °C overnight. The 5 ml were transferred into 200 ml of LB medium with antibiotic and cultured at 200 rpm at 37 °C overnight. Bacterial cells were collected by centrifugation at 4000 rpm for 20 min at 4 °C. Supernatant was removed and cells resuspended in 50 ml cold GET solution. 50 ml of NaOH/SDS solution (room temperature) was added to lyse cells and samples were kept at room temperature for 5 min before adding 50 ml of cold potassium acetate solution. In this step proteins are precipitated in theory trapping long bacterial chromosomes, so that smaller DNA constructs such as fosmids are still in solution. Samples were centrifuged at 4500 rpm for 40 min at 4 °C and supernatant was transferred into a new tube by filtering through Propax gauze (Smith and Nephew, U.K.). After precipitating DNA by adding isopropanol (0.7vol) and centrifuging the samples at 4500 rpm for 40 min at 4 °C, the DNA was washed once with 25 ml 70% ethanol and centrifuged again at 4500 rpm for 30 min at 4 °C. The supernatant was discarded and the DNA pellet air dried for 30 min. DNA was dissolved in 0.5 – 1 ml of TE buffer. DNA was quantitated using the Nanodrop® fluorometer and directly used for labelling in the FISH experiments which were carried out by Jill Brown and Veronica Buckle (Weatherall Institute of Molecular Medicine, Oxford, U.K.).

7.6.2 2D and 3D FISH experiments

Fluorescence in situ hybridisation was carried out in cooperation with Jill Brown and Veronica Buckle at the Weatherall Institute of Molecular Medicine, Oxford, U.K., using the fosmids described above.

Briefly, fosmids spanning the region of interest were amplified, isolated and labelled by nick translation with biotin or digoxigenin. 2D FISH was performed as by Brown et al. (Brown, Saracoglu et al. 2001) except that slides were denatured at 72 ^{o}C and probes were denatured at 85 ^{o}C for 10 min. For each probe pair, inter-probe distances were measured in micrometers using ImageJ software in approximately 100 nuclei. 3D FISH was also performed as described in Brown et al. except that cells were permeabilised in 0.5% Triton X-100 in PBS followed by quenching in 0.276% ammonium chloride in PBS for 10 min at RT. Cells were denatured in a 1 in 4.5 dilution of conc. HCl in H_2O. Slides were hybridised overnight at 37 ^{o}C. Images were captured on a Zeiss 510 confocal microscope (Zeiss, Germany). Z stacks were taken of each nucleus at 400 nm steps. For each probe pair, inter-probe distances were measured in micrometers using Zeiss LSM software and nuclear volume measurements in ImageJ software using the Voxel Counter plug-in in approximately 100 nuclei. Non-parametric statistical analysis was by Kruskal-Wallis 1-way ANOVA and by Mann-Whitney U-test (SPSS 19).

7.6.3 Chromatin immunoprecipitation (ChIP)

The protocol used for ChIP was modified from the Upstate protocol (Figure M3). Cells were grown in 15 cm dishes to sub-confluency, exposed to normoxia, hypoxia (0.5-1%) or dimethyloxalylyglycine (DMOG, 2 mM) for 16 hr.

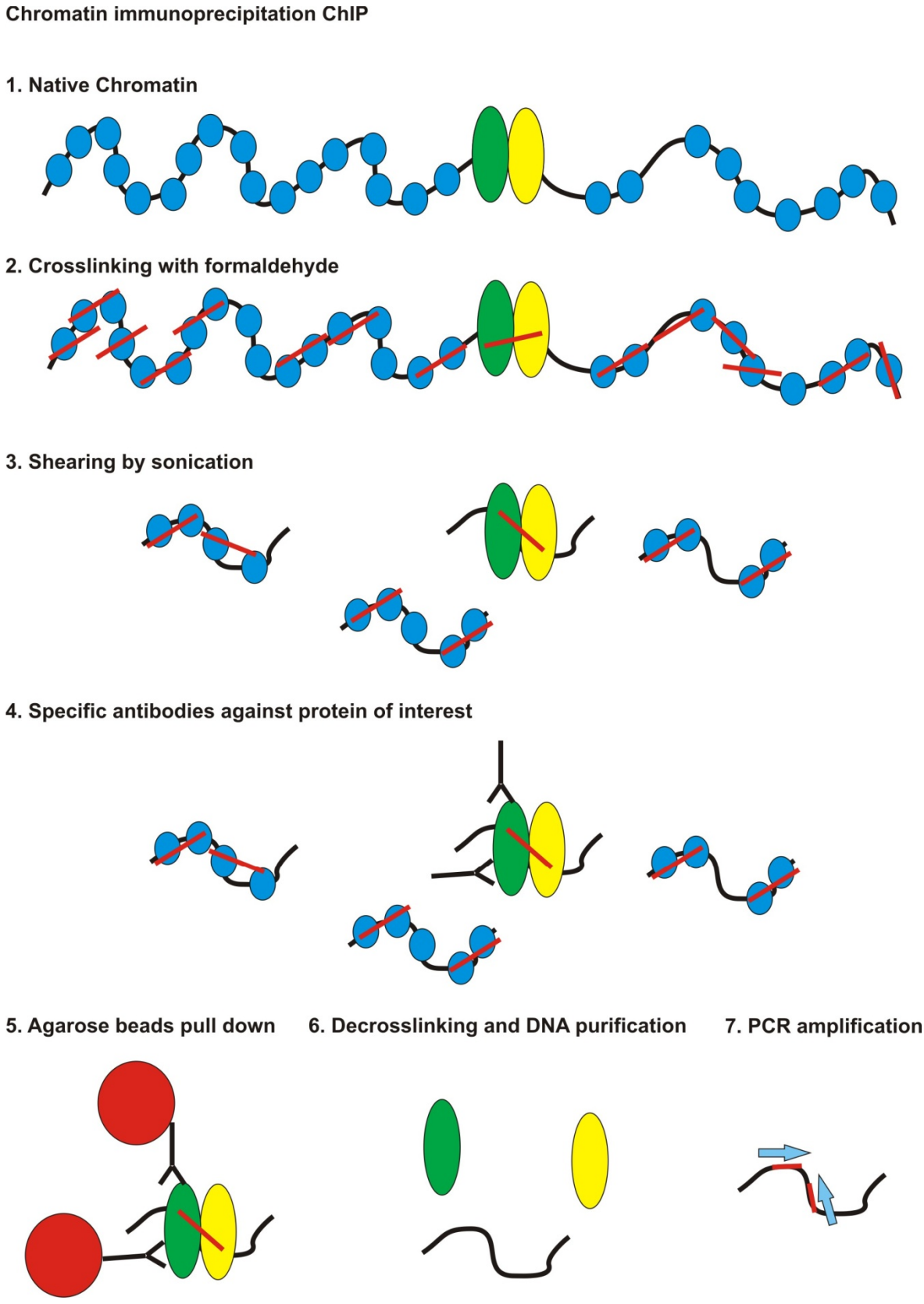


Figure M3: Schematic overview of the chromatin immunoprecipitation technique.

1. Cells were cross linked using 1% formaldehyde for 12 min on ice.
2. Cross-linking was stopped by adding 1 M glycine to a final concentration of 125 mM for additional 5 min.
3. Cells were washed three times with ice-cold PBS, scraped in 5 ml of ice-cold PBS and transferred into a 15 ml tube (TPX tubes, Diagenode, U.K.).
4. Cells were pelleted at 1000 rpm at 4 °C for 5 min, the supernatant was removed and the pellet washed with 5 ml ice-cold PBS.
5. After recentrifugation and removal of the supernatant the cells were lysed in 1 ml ChIP lysis buffer for 10 min on ice.
6. The cell lysate was sonicated using a water bath sonicator (Bioruptor, Diagenode, U.K.) filled with iced water for 2 x 7.5 min in pulses of 15 sec on and 15 sec off. In between the two cycles ice was replaced to ensure sufficient cooling of the samples.
7. The lysate was transferred into a 1.5 ml tube and centrifuged at 13000 rpm at 4 °C for 10 min to remove cell debris. The sonicated chromatin was then transferred into 15 ml tubes and diluted with ChIP dilution buffer (1 in 10).
8. Using 150 µl of 50% Protein A agarose (Millipore, U.K.) slurry, which was washed three times in dilution buffer, the lysates were “pre-cleared” at 4 °C for 60 min on an end-over-end rotator.
9. Samples were centrifuged at 1000 rpm for 5 min at 4 °C and 2-4 ml aliquots of the supernatant were transferred into fresh 15 ml tubes for immunoprecipitation.
10. 100 µl of diluted chromatin were saved in a 1.5 ml tube as input sample.
11. To each aliquot 5-10 µl of specific antibody was added including appropriate control antibodies or sera.

12. Samples were incubated overnight at 4 °C on an end-over-end rotator.
13. 60-80 µl 50% Proteinase A agarose beads slurry was added and samples were rotated for another 60 min.
14. After centrifugation at 1000 rpm a 4 °C for 5 min the supernatant was removed, samples were resuspended in 1 ml of ice-cold low salt washing buffer and transferred to new 1.5 ml tubes.
15. Samples were incubated at 4 °C for 5 min on an end-over-end rotator and then centrifuged at 2000 rpm at 4 °C for 3 min.
16. Supernatant was removed and washing steps were repeated using 1 ml of ice-cold high salt buffer, LiCl buffer and TE buffer (twice).
17. TE buffer was carefully removed and chromatin was eluted from the beads by adding 250 µl elution buffer and rotating end-over-end at room temperature for 15 min.
18. Samples were centrifuged at 2000 rpm at room temperature for 3 min and the eluate was transferred into a fresh tube.
19. The elution step was repeated with a further 250 µl of elution buffer.
20. 20 µl 5 M NaCl were added to the 500 µl sample (including the 100 µl input sample diluted with 400 µl elution buffer).
21. Samples were incubated at 65 °C overnight to decrosslink the chromatin.
22. The next day 10 µl of 0.5 M EDTA, 20 µl of 1 M TRIS pH6.5 and 2 µl of 20 mg/ml proteinase (Roche, U.K.) were added and samples were incubated at 45 °C for 60 min.
23. Samples were transferred to fresh 1.5 ml tubes and DNA was extracted using phenol chloroform (see above).

24. DNA from the ChIP experiments was resuspended in 25 µl H_2O.
25. Aliquots of ChIP DNA samples were further diluted 1 in 10 in H_2O for qPCR analysis and 2 µl of this dilution was used per ChIP qPCR reaction.

HIF enrichment levels of the specific DNA interaction proteins were determined by qPCR using the ΔΔct method. Values were normalised to an internal control (beta actin) and related to values from the negative control antiserum. For histone modifications values were normalised to total H3 levels present at the DNA site of interest and were calculated as percent of input.

7.6.3.1 Quality control for ChIP

Sonication efficiency was verified by running input DNA samples on an agarose gel. Usually the average fragment size was about 300-500 bp. Enrichment of HIF, RNA pol2 and histone marks at control loci was confirmed by qPCR. The *EGLN3* intronic enhancer or an intragenic site of beta actin served as positive controls for HIF and RNA Pol2, respectively. For histone methylation sites positive controls were generated using the data set from Barski et al. (Barski, Cuddapah et al. 2007).

7.6.3.2 ChIP-sequencing

ChIP-sequencing was carried out in cooperation with Lorna Gregory, Spyros Oikonomopoulos and Jiannis Ragoussis from the Genomic sequencing facility at the Wellcome Trust Centre (University of Oxford, U.K.) (Figure M4). For ChIP-seq experiments an aliquot (15 µl of 25 µl total) of captured DNA fragments dissolved in water was forwarded to the Genomic sequencing facility. Libraries for sequencing were prepared using the Illumina ChIP-Seq kit. 36, 52, or 76 bp single end sequence analysis was

performed on the Illumina GAII platform. Derived sequences were mapped to NCBI build 36 (Human Genome 18, hg18) using the ELAND software package (Illumina, U.K.) or BOWTIE mapping software (http://bowtie-bio.sourceforge.net/index.shtml). Subsequent assessment, peak finding and motif analysis and mapping of RCGTG motifs were undertaken using the CisGenome software suite (http://www.biostat.jhsph.edu/~hji/cisgenome). The Cisgenome software allows the direct comparison of two samples (e.g. HIF ChIP vs preimmune control ChIP) to identify significant peaks of binding and to correct for background signals. In addition to the number of reads per window, three additional filters were applied to define high stringency peaks. Firstly, the number of reads in each direction must comprise at least one third of the total (i.e. the peak height for both the Watson and the Crick frames must be greater than 7). Secondly, the estimated false discovery rate must be less than 0.05. Finally gene loci with more than 4 binding sites were excluded as manual analysis revealed these to be due to regions of high background, mapping to repeat elements or contamination with expression plasmids.

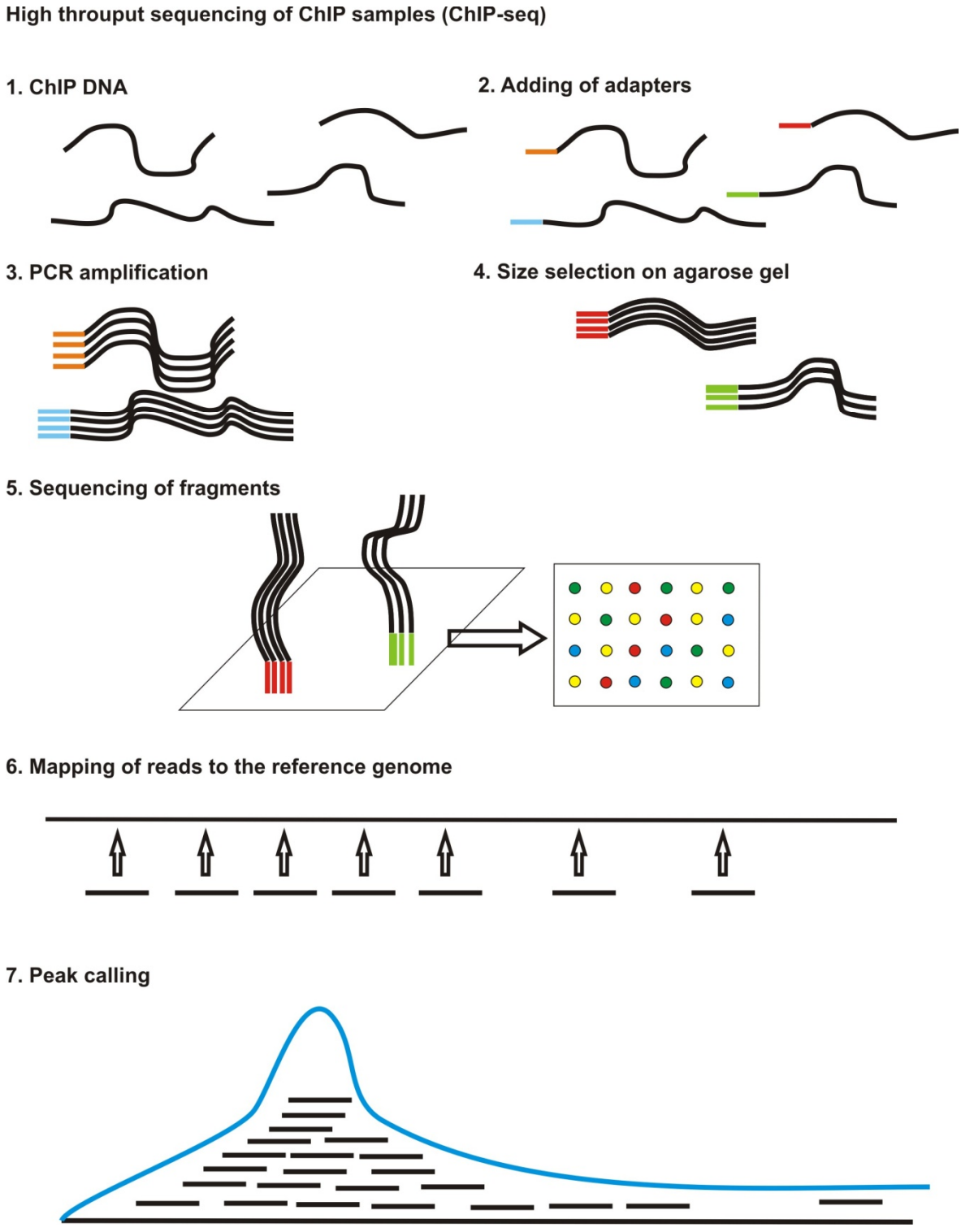


Figure M4: Schematic overview of the ChIP-seq workflow.

7.6.3.3 Data and Data analysis tools

MCF-7 expression data were previously published and are available at

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE3188.

ENCODE Digital DNAse1 hypersensitivity data were downloaded from:

http://genome.ucsc.edu/cgi-bin/hgTables.

Data analysis was performed using publically available academic software packages and web resources:

Cisgenome - http://www.biostat.jhsph.edu/~hji/cisgenome/

GSEA - http://www.broadinstitute.org/gsea/index.jsp

DAVID - http://david.abcc.ncifcrf.gov/

Galaxy - https://main.g2.bx.psu.edu/

UCSC Genome Browser - http://genome.ucsc.edu/

GeneVenn - http://genevenn.sourceforge.net/

ChIP-seq data were deposited at the GEO database online: GSE28352 (MCF-7 HIF-ChIP) and GSE34871 (786-O HIF-ChIP).

7.6.3.4 Independent confirmation of enrichment by real-time qPCR

A random selection of peaks identified by ChIP-seq was confirmed by SYBR green® based real-time quantitative PCR (Figures 2.5 and 4.6) using a StepOne thermocycler (Applied Bioscience, U.K.). For details on the qPCR set-up please see the qPCR section below. β-actin DNA was used for normalization and fold-enrichment at each site was determined using the $\Delta\Delta C_T$ method.

7.6.4 Formaldehyde assisted isolation of regulatory elements (FAIRE)

Following the protocol from Giresi et al. two aliquots (50 µl each for FAIRE and input control) from cross linked and sonicated ChIP chromatin were used (Giresi, Kim et al. 2007; Giresi and Lieb 2009). The principle of this method is that DNA in regulatory regions is depleted from histone proteins (Figure M5). Interactions with transcription factors are weaker than with histone proteins. Therefore, in a phenol chloroform extraction, which is used to separate proteins from DNA, sonicated DNA cross linked with proteins will remain in the protein phase, whereas DNA not bound to proteins (e.g. in regulatory regions) will appear in the aqueous phase. This step will enrich for naked DNA, i.e. regulatory elements, which can be quantified by qPCR with reference to the input of total DNA from this region.

1. For the FAIRE assay one 50 µl aliquot was diluted with 450 µl of TE-buffer and two rounds of phenol-chloroform extraction (see ChIP protocol) were performed followed by de-cross linking by heat.
2. For the 50 µl input control de-cross linking by heat was performed directly without previous phenol-chloroform extraction.
3. After Proteinase-K digest (see ChIP protocol) DNA from both de-cross linked samples was purified by phenol-chloroform extraction as outlined in the ChIP assay followed by a column-based DNA "clean-up" (Sigma, U.K.).
4. DNA concentrations for all samples were adjusted to 2 ng/µl and qPCR with primers of interest was performed in triplicates with input and FAIRE DNA.
5. Values were normalised to input DNA and compared to a region just outside of the putative regulatory regions (*NDRG1* control region human genome 18 (HG18) chr8:134458206-134458299, *CCND1* control region HG18 chr11:68948541-

68948669). Fold enrichment is expressed compared to the negative control region.

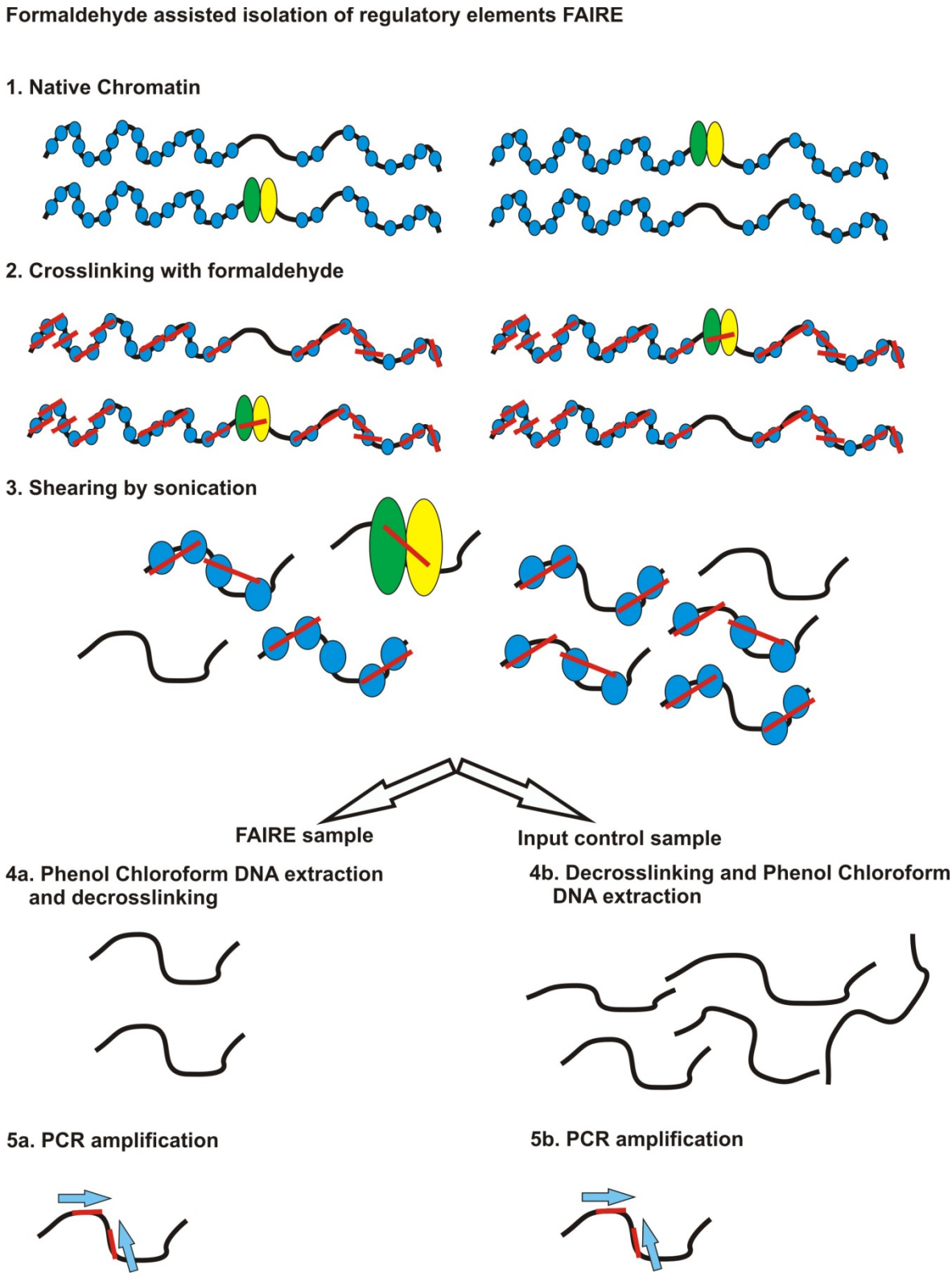


Figure M5: Schematic overview of the FAIRE assay.

7.6.5 Chromatin conformation capture (3C) assay

Experiments were performed according to the protocol of Hagege et al. with some modifications (Figure M6) (Hagege, Klous et al. 2007).

1. Cells were grown in 15 cm dishes with (MCF-7) or without (786-O, UMRC3) 1 mM DMOG.
2. One 15 cm dish of cells per cell line was used and cells were cross linked using 1.5 mM Di(N-succinimidyl) glutarate in 10 ml PBS for 45 min at room temperature followed by three wash steps with 20 ml PBS and further cross-linking with 1% formaldehyde in PBS for 10 min at room temperature.
3. Reactions were quenched by adding ice-cold glycine to a final concentration of 0.125 M and kept on ice for another 5 min at room temperature.
4. Cells were scrapped in 5 ml ice-cold PBS per plate and transferred to a 15 ml tube. Cells were pelleted at 500 rpm at 4 °C for 5 min and washed with 5 ml ice-cold PBS.
5. After pelleting cells were lysed for 30 min in 5 ml 3C lysis buffer on ice resuspending the cells every 5 min.
6. Integrity of nuclei was checked by microscopy and nuclei were counted. $1x10^7$ nuclei were washed and resuspended in 0.5 ml 1.2x EcoRI restriction enzyme buffer (Fermentas, U.K.).
7. 7.5 µl of 20% SDS (final 0.3%) were added to the nuclei suspension and samples were incubated at 900 rpm at 37 °C for 1 h.
8. 50 µl 20% Triton X-100 (final concentration 2%) were added and nuclei incubated at 900 rpm at 37 °C for a further 1 h. At this stage a 5 µl aliquot was taken as an undigested genomic DNA control.

9. Restriction enzyme digest was carried out by adding 600 U of EcoRI (Fermentas, U.K.) and incubating the samples at 900 rpm (Thermomixer, Eppendorf, Germany) at 37 °C over night.
10. The next day 40 µl of 20% SDS was added and samples incubated at 900 rpm at 65 °C for 20 min.
11. Samples were transferred to a 50 ml tube and 6.125 ml of 1.15X ligation buffer (Fermentas) was added to dilute the fragments. After adding 375 µl of 20% Triton X-100 samples were incubated at 37 °C for 1 h while gently shaking.
12. 10 µl of T4-DNA ligase (Fermentas) were added and samples incubated at 16 °C for 4 h followed by room temperature for 30 min.
13. 30 µl of Proteinase-K (10 mg/ml) were added and samples were incubated at 65 °C overnight to de-cross link samples and remove proteins.
14. The next day 30 µl of RNAse (10 mg/ml, Sigma, U.K.) was added and samples incubated at 37 °C for 45 min.
15. DNA from the 3C samples and the de-cross linked non-digested samples were then isolated using phenol chloroform extraction.

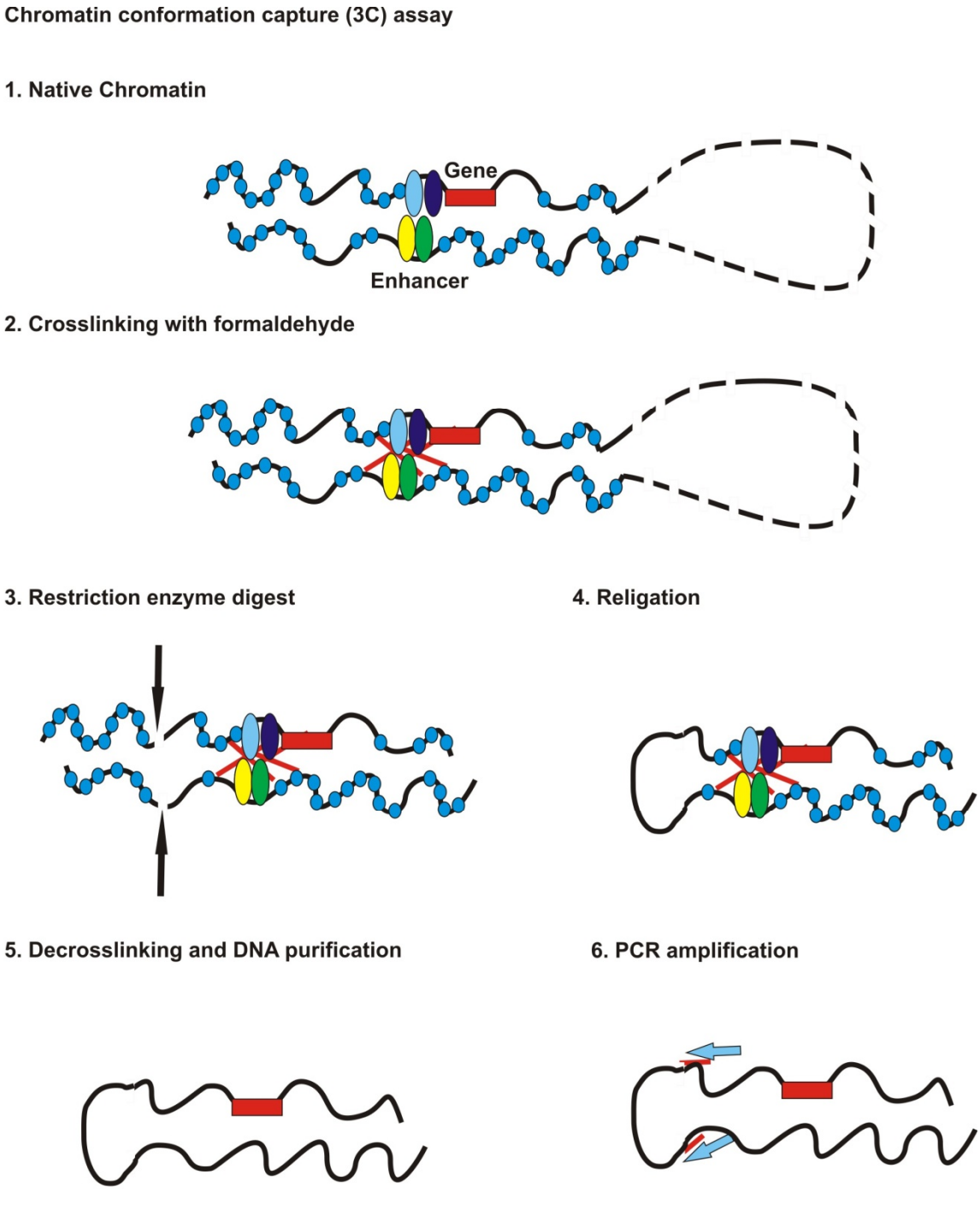


Figure M6: Schematic overview of the chromatin conformation capture (3C) assay

7.6.5.1 Quality control for 3C assays

The quantity and purity of DNA was determined using a NanoDrop® ND-1000 spectrophotometer (Thermo Scientific, U.K.). Digest efficiency was monitored in qPCR experiments using primers spanning restriction sites and comparing digested to non-digested DNA (Figure M7).

7.6.5.2 qPCR analysis of 3C templates

Chromatin interaction was interrogated using quantitative PCR employing a forward primer and a Taqman® probe located just upstream of the first EcoRI restriction site downstream of the 11q13.3 HIF-binding locus (the bait) combined with a series of reverse primers covering consecutive EcoRI restriction sites across the region of interest. Two BAC clones covering the HIF-binding site and the *CCND1* gene, respectively, were used to create an artificial library of ligation products in order to normalise for PCR efficiency (Figure M7). On every PCR plate BAC clone DNA was run as a control for the specific products. Data were normalized to the signal from the BAC clone library and between cell lines by reference to a region at the *EEF1G* locus.

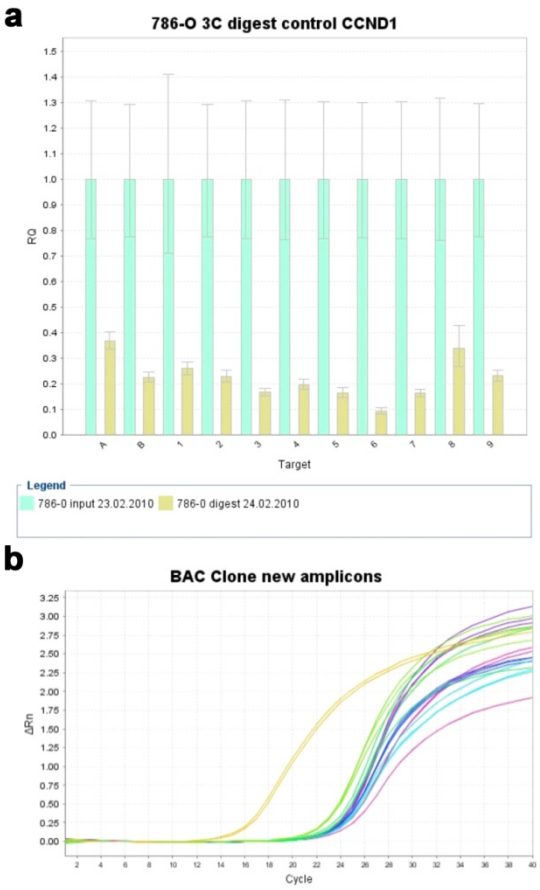


Figure M7: (a) qPCR analysis for different primer pairs spanning EcoRI restriction sites before (green) and after (brown) EcoRI digest. Values were normalised to signals from an amplicon not spanning a restriction site and to input values. Restriction digest efficiency should be 60% or higher. (b) Amplification curves using primers as in (a) of newly ligated fragments from digested BAC clones. The brown curve indicates an amplicon not spanning an EcoRI restriction site. The curves indicate equal re-ligation efficiency across the different EcoRI restriction sites tested.

7.6.6 Allele specific assays

ChIP DNA, FAIRE DNA and cDNA were prepared as above and submitted to Chiara Bardella and Ian Tomlinson (Wellcome Trust Centre for Human Genetics, University of Oxford, U.K.) for allele specific assays. Linkage disequilibrium of SNPs in vicinity of the HIF-binding peaks (rs4980785, rs7948643, rs7939721, rs7939830, rs17136556, rs77247065, rs11263441 with the tag SNP rs7105934 on 11q13.3 was verified by them in a cohort of 192 cancer-free control individuals using competitive allele-specific PCR KASPar chemistry (KBiosciences Ltd, Hertfordshire, U.K.). Haploview software (www.broadinstitute.org) was used to define the haplotype blocks and recombination hotspots. To identify samples

heterozygous for the common and rare alleles at rs7105934, rs7948643 and rs77247065, pVHL-defective ccRCC cell lines were genotyped using KASPar. For the heterozygous (1:1 allelic dosage) cell line KTCL140, genomic DNA derived from ChIP or FAIRE (two to three independent biological samples) or cDNA generated from DNAse-treated mRNA were genotyped at rs7948643 and the *CCND1* 3'UTR SNPs rs7177 and rs678653, using allele-specific Taqman® assay (Applied Biosystems, U.K.). Constitutional DNA from individuals known to be homozygous or heterozygous at each SNP was included in each assay as a reference. The KTCL140 cell line DNA was also genotyped using the Illumina OmniExpress SNP array according to the manufacturer's protocols. The data analysis was performed using Illumina Genomestudio software.

7.7 Antibodies

Name	**Source**	**Cat nr.**	**Host**	**Application**
HIF-1β	Novus biological	NB 100-110	Rabbit antiserum	ChIP, Western
HIF-1α	Laboratory	PM-14	Rabbit antiserum	ChIP
HIF-2α	Laboratory	PM-9	Rabbit antiserum	ChIP
Preimmune	Laboratory	PM-14 pre	Rabbit serum	ChIP
HIF-1α	BD Bioscience	610959	Mouse monoclonal	Western
HIF-2α	Laboratory	190b	Mouse monoclonal	Western
H3	Abcam	ab1791	Rabbit polyclonal	ChIP, Western
H3K4me1	Abcam	ab8895	Rabbit polyclonal	ChIP, Western
H3K4me2	Abcam	ab7766	Rabbit polyclonal	ChIP, Western
H3K4me3	Cell signalling	9751	Rabbit monoclonal	ChIP, Western
H3K9me1	Upstate	07-436	Rabbit polyclonal	Western
H3K9me2	Cell signalling	9753	Rabbit polyclonal	Western
H3K9me3	Upstate	07-523	Rabbit polyclonal	Western

H3K27me1	Upstate	07-448	Rabbit polyclonal	Western
H3K27me2	Cell signalling	9755	Rabbit polyclonal	Western
H3K27me3	Upstate	07-449	Rabbit polyclonal	Western
H3K36me1	Abcam	ab9048	Rabbit polyclonal	Western
H3K36me2	Upstate	07-369	Rabbit polyclonal	Western
H3K36me3	Abcam	ab9050	Rabbit polyclonal	Western
H3K79me1	Abcam	ab2886	Rabbit polyclonal	Western
H3K79me2	Cell signalling	9757	Rabbit polyclonal	Western
H3K79me3	Abcam	ab2621	Rabbit polyclonal	Western
H4K20me1	Abcam	ab9051	Rabbit polyclonal	Western
H4K20me2	Cell signalling	07-031	Rabbit polyclonal	Western
H4K20me3	Upstate	07-463	Rabbit polyclonal	Western
H3K9ac	Upstate	06-942	Rabbit polyclonal	Western
H3K14ac	Upstate	06-911	Rabbit polyclonal	Western
H3K27ac	Abcam	ab4729	Rabbit polyclonal	ChIP
RNA pol2	Santa Cruz	Sc-899	Rabbit polyclonal	ChIP
Glut1	Abcam	ab652	Rabbit polyclonal	Western
Ca9	Abcam	ab15086	Rabbit polyclonal	Western
Rabbit IgG	Millipore	12-370	Rabbit IgG	ChIP

7.8 Primers and Taqman® assays

7.8.1 Primers

Binding region ID	Forward primer	Reverse primer	Application
beta actin	ACCATGGATGATGATATCGCC	GCCTTGCACATGCCGG	ChIP-seq validation
EGLN3 positive control	AGTGTCCGTTCCCAGCTCAG	TAGGCACAGTAAACAGGCC	ChIP-seq validation
PKM2	TATGGACAGAAGTGGGGGAG	CTGCAGTGGTTTGTGCTTGT	ChIP-seq validation
SLC2A3	TGTGTATGCGTGAGTCTTTGG	GAGCCCCGAGCTACATTCCT	ChIP-seq validation

ZFAT1	CTCGCTTCCCATTTCCTGT	GTGTGAGTGAGCACGTGAG	ChIP-seq validation
P4HA1	GAGCGGCTACTTCCTACCCT	ATAAAAGGGCTAACGGGCT	ChIP-seq validation
DNAJB6	GGCGATGGCAGAGATAAGAA	ACGCAGCTATTGCACAACA	ChIP-seq validation
SLC28A1	GGATCCACTGGACAAGCAGT	CACTCCTGGGACTGAGATGC	ChIP-seq validation
NUPL1	AGTACGCCTATTCTCGCCTG	AATGAGCCCTGGTACGTGA	ChIP-seq validation
PNPO	GTGAACTGTGCTTCCTTCCC	GGGATAAACAAGGAGAGGG	ChIP-seq validation
SOS1	CTTCTTTTCGCTTCCCTCCT	CCTCTCGGTCTTCGCCTATT	ChIP-seq validation
UTRN	TCAGATGAAAGATACGGGGC	TAAAAAGGCCTCTCAGCAG	ChIP-seq validation
FAM36A	TTGTTTTCCGAACCCATCTC	CGAACGGGTCACACTTTCTT	ChIP-seq validation
MGAT5	TCTCTTCCCTGAGGCTTGTC	CTGGATCTGTGCATGGATGA	ChIP-seq validation
AKR1CL2	TGTTTGCTTTGACTACATAATT	TTGTGAATCACCTCTGCAAT	ChIP-seq validation
CUTL1	GCAGCTGGACTCATCTCACA	TTAAAAACACGTGCCTTCTG	ChIP-seq validation
BBX	CGAGCTCTGCCTCTGCTAAT	TAGCAGGACAGGCAATGGT	ChIP-seq validation
DPT	ACTGAGAGACTGGGTGTGGG	AAGTGGTGTGCGTGGAAAC	ChIP-seq validation
S100P	GACACCACGTTTTACTGCCC	GCCACAACAACAAAGCTGT	ChIP-seq validation
AK3L2	CTCCTCTCCACCGCTCCT	GGGTTTTCCCTGCCTAGTTC	ChIP-seq validation
PGM1	GCGGAGAACTTCATCCAGAG	GGCCTCCTTCATGTAGAACC	ChIP-seq validation
PPFIA4	GCTCGGCTCTAGGACGTGT	GCTCACCTGCCTGCAAAT	ChIP-seq validation
PDK1	CACGTCCCTCACGTACCAC	ATGCCGAGTCCCGCTAGA	ChIP-seq validation
BHLHB2	CTGGAGCTAGCAAGGGGATA	GGAGGGGGAGAAGTTGGTAA	ChIP-seq validation
STBD1	GGGTAGGTGTGGCTCTGTGT	CAGCCAAACGAAAAGTGCTC	ChIP-seq validation
P4HA2	AGCCCACCGCACGTACTC	GTGGCGACCTCTCCAGAC	ChIP-seq validation
SPOCK1	TCTGGGGCCTAAAGCCTATT	GCTCAGGATGCCAATTTGTT	ChIP-seq validation
GJA1	TCAAGCTAAGGGTGGCTGTT	ATAACCTCTGCCCCTCCAGT	ChIP-seq validation
VEGFA	CTCAGCCTGTTCAGCAGTTG	AGACTCAACGTGCCAGGAGT	ChIP-seq validation
ZNF292	CCCTTTCCTTCCCTGAACTC	AAGAGAACCTGGCCTCCATT	ChIP-seq validation
EGFR	CTCCAGGAAGCCTACGTGAT	GTCTTTGTGTTCCCGGACAT	ChIP-seq validation
IGFBP3	CGGTGACAGTGTGGTTGTCT	CCCGCTTGCTTAACTCTCAC	ChIP-seq validation
NDRG1	TCTTCCTCCCACATGGAGAC	CACGTGTTGCTGAGTCACCT	ChIP-seq validation
ZNF395	AGGAACTGGAACAGGCAGAA	AGACATGTGGTTGTCCAGCA	ChIP-seq validation
ADM	CTCCGGAAAACCCAGTAACC	TAAGGCTGTGTGCTGACTCC	ChIP-seq validation

FOS	AAATGCCCTTCCCACCTACT	AGGGAGGGCACTACCTGAGT	ChIP-seq validation
SCNN1B	CTCGCCCTAGGACACATAGC	GCACAGCTGATGACAGAGGA	ChIP-seq validation
SOX9	CTCTGTCAGGCCCTAACTCG	TTCCCCACACTTCAGTCACA	ChIP-seq validation
ANGPTL4	TCTCCACCTGCCTGAGAACT	GCAGTGAGCTGAGATTGCAC	ChIP-seq validation
GYS1	GGCCTTCAGCTCTGTCAATC	AATGTCTCCTGGCTCCTCCT	ChIP-seq validation
EGLN1	GGAGCAGCCACAAAGAAAAG	TATTGCGGTTTTTCCTCCAC	ChIP-seq validation
C6orf206	GGGTGAAGGTGACAGCAAGT	ACACAGACTCAACGTGCCAG	ChIP-seq validation
SFRS3	AGGCCTGTTACAACGAGCG	CTACAAACGCAGCGAGAGC	ChIP-seq validation
ZC3HAV1L	GGTGAGGAAGGAGCACACTG	GTGACTGGGAGGAGCCTG	ChIP-seq validation
SHMT2	CTCTCAGAGCCTGGGCTAAC	TACGTGAGCCAACAAGAAGC	ChIP-seq validation
CDH2	AGACGGGGGTCACGTTTC	TACCCACAAACCCGCATACT	ChIP-seq validation
DBC1	CGTGAGGGCTGTCATGTTTA	GAAAATAAGGGCAAGCGATG	ChIP-seq validation
FGD5	CATCGCAGAGTCACTGAAGC	CAGACGTGCAGATTCTGGGT	ChIP-seq validation
JUN	GCCCAGAATTATGGAGCCTA	GTCTGGATGCAGCCACTGT	ChIP-seq validation
SEPT9	ACGGAAGCCTGTCATTTCC	TAGCTCATGTCTGTGCCTCG	ChIP-seq validation
CLINT1	TGTTCCTGACAATGCAGCTC	AGCCCACACACTCTAACCCA	ChIP-seq validation
STC2	GACAGACGTGTTCTTGCTCG	CTCAGCACCCACGAACTATG	ChIP-seq validation
ROR2	AAAGAACAAGTGCTGGTGGC	TGACCCTTGAGGTTTTGCTC	ChIP-seq validation
MYH13	ACCACTGGTATGATTTGGGC	GTTGCACATCCCACATTTGA	ChIP-seq validation
TMTC2	CTTTCTTTCCCTTTTTGGGG	CTACGGTTGGCCACAATGTC	ChIP-chip validation
PGK1	CGTTTTGTGAGCGTTGGTTA	ACTCTGGCCCCCACCTTT	ChIP-chip validation
CA9	ACAATGTGCATATCGTGGCA	CAGAGGGGCTCTGATAGGTTT	ChIP-chip validation
YEATS2	GAGGGAATGGTGGGAATCTT	TTCGTTTTTCCCACCTCAAA	ChIP-chip validation
SPRY1	GAAAAAGCAGGGTCCCTTTC	GGCATTAACGGCTCTGTGTT	ChIP-chip validation
DTNA	TTTTCAATGACGAGCGTGC	CTTCCTTCCTGGACCGAAAT	ChIP-chip validation
RAD51L1	TTTTCTTTTCTTGGTGGTAAGGA	TGTGTATTGCACATGAGGGG	ChIP-chip validation
SLC2A1	CTGATTTGCACATCTGCTTGA	GTCGGAGGGCATACAGAAGA	ChIP-chip validation
GPRC5A	TGGTTAGCTGGGAGGTATGG	GGGGCACGTCTATTATCAGC	ChIP-chip validation
HIG2	GGGGAGAAGGGATATGTGCT	CGACTGCCATCTCTGCTGTA	ChIP-chip validation
HK2	ATGGGCCGACTCTTGTATTG	GGAAAGTGAGCCGCCTTAG	ChIP-chip validation
S100P	TGTGGCCGTTTCTTGAAAAT	AAGCTGTCTCCACGTCCACT	ChIP-chip validation

NARF	CTTATGTCACACTCTCCCAGACA	GGGCTATAATGGAAGTGTTGGA	ChIP-chip validation
C3ORF28	CAGATTAGCCTGAGGGATGC	CATAATTGGACTCTGCTCAGTGG	ChIP-chip validation
ARRDC3	CCCCTGCAGTCACACACTC	TTTGTCACATGGGACTCTTC	ChIP-qPCR
ERRFI1HIFbind 1	GATTACAGGCTGGATGGCAC	TGCTGCCAGACTGGTATGAG	ChIP-qPCR
ERRFI1HIFbind 2	CAAGTGTGCACGTTTGAACC	ATCTGTGGAAAGGCAGGAG	ChIP-qPCR
MLL5 HIF bind	AAACCCCACAGGAATGCAG	ATGGCGCTGGGTCTCTTT	ChIP-qPCR
SAP30 HIF bind	TCATATTAGCCACGCCCATC	CGTACGCGTGTATGTTCCTG	ChIP-qPCR
CCND1 second HIF binding site	CCACTGTCCAGACCCTCTGT	GCAGGTCAGGACACAGTTCA	ChIP-qPCR
CCND1 HIF bind 3rd peak	ACAGGGGCAGACAGGTGTAG	GTGGCTGACCACATTCCTG	ChIP-qPCR
ALDOC promoter	TCTGCCACAATCCCTTTACC	TGCCTCCTCTTCACCAGACT	ChIP-qPCR
FAM13A1 promoter	TGCTCTCAGCCTCCAAACTT	CCCTGGAACAAAAGGTCAAA	ChIP-qPCR
HIF1A antisense promoter	GCTACATGATGAGGATGCGA	CCGAGCACGTACTGAGGC	ChIP-qPCR
NICl promoter	TGTGTATGCGTGAGTCTTTGG	GAGCCCCGAGCTACATTCCT	ChIP-qPCR
NDRG1HIF bind	CCAGAGGTGACTCAGCAACA	CTGCATCTGGCAAGAGGTG	ChIP-qPCR
CCND1 enhancer negative control	GGATCACCAGGTGTATTCGG	CATCCAAGGGGGATACACAC	ChIP-qPCR
hu CCND1 HIF bind 2	GTACGTGTCCCAGGCAGC	TCAAGGACCTGAGGTTGTCA	ChIP-qPCR /FAIRE
CCND1 HIFbind4fw-rev	CAGTCACGGACACTGAGGTG	AGAGGCTCCCTGTCATCAAG	ChIP-qPCR /FAIRE
CCND1 HIFbind3fw-rev	CTTGATGACAGGGAGCCTCT	CCGGTGCTCAGCAAACTTAT	ChIP-qPCR /FAIRE
ARRDC3HIFshortB	TAGGATCCACAGGAGTAGCGCGAGCAG	TAGGATCCGAATCCTGGATGTCCCCAG	Cloning
ERRFI1HIF1shortB	TAGGATCCCATGTGACGTCAGAAAGGCT	TAGGATCCCTGCAAAGACCTTTGAGTTCC	Cloning
CCND1 Pr K	ATGGTACCACCCCCAACAAAACCAATTA		Cloning
CCND1 Pr N	AGCTAGCACTCCCCTGTAGTCCGTGTG		Cloning
Nici Pr N	AGCTAGCTCCCTTTCTTGCACTCGTCT		Cloning
Nici Pr K	ATGGTACCAAAAGACTCAGCGTGCTGCT		Cloning
CCND1 En shorter B fw	ACGGGATCCCCGTGGCAACATTACATCAC	ACGGGATCCAGAGGCTCCCTGTCATCAAG	Cloning
CCND1 En B rev long		ACTGGATCCTTCATGGAGGAGAAACTGGG	Cloning
CCND1 Pr K fw	ATGGTACCACCCCCAACAAAACCAATTA	AGCTAGCACTCCCCTGTAGTCCGTGTG	Cloning
CCND1 En 3 K fw	ATTAGGTACCCCTCTCCAGGTGAACGTTGT	ATTAGCTAGCGTGTTCTCCGTCTCTGCCTC	Cloning
Minor peak h2 HRE1	CATGAACCAGCTCACATACCTGGCCCAC	GCCAGTGGGCCAGGTATGTGAGCTGGTTCA	SDMG

	TGGC	TG	
Minor peak h2 HRE2	CACCCCAGCCAGGCATACTGTGTCAGCCC	GGGCTGACACAGTATGCCTGGCTGGGGTG	SDMG
Major peak H	CTGGGCAGCCGTACATATCCCAGGCAGCGG	CCGCTGCCTGGGATATGTACGGCTGCCCAG	SDMG
rs7948643	GCTCAGGATCACCGGGCAGCCGTAC	GTACGGCTGCCCGGTGATCCTGAGC	SDMG
rs7939721	CAGCGGGAGCCCTGGCAGTGCCAGC	GCTGGCACTGCCAGGGCTCCCGCTG	SDMG
Taqman® assay: CCND1 ECORI_1	CCACTGGACCCTTGAATCTC	CTTGCCACATTGCTTTCAGA	3C assay
Taqman® assay: CCND1 ECORI_1 probe	CTGAACGAAGCCCACCATCAGA		3C assay
ECORI site HIF binding A-8	TGTCTGCATGGACACAGCC	ATTCTTTCTGACCACAGGGC	3C assay
ECORI site HIF binding A-4	GTGTGCTGATGTTTTGCCAT	ACTGCACCTGGCCTCATAAT	3C assay
ECORI site HIF binding A-3	TGTCTTCTGCACTTTCCTGG	AAGAAAGCCCTGGACCTGAT	3C assay
ECORI site HIF binding A-2	AAAGCGAAGTCCCAGTCTCA	TGGTTGAAGCACACAGTGGT	3C assay
ECORI site HIF binding A-1	TTGAAGAGGATGAGCAAGGG	CGAGTTTTAAGTGATTTGCCC	3C assay
ECORI site HIF binding B-1	CTGGCTCCGACTTCCAGG	TTCATGAACCAGCAACCTCC	3C assay
ECORI site HIF binding B-3	CTCAACCCACCTTGCAAAAG	CCCCCTACTCGGGAAAATTA	3C assay
ECORI site HIF binding B-4	TGCTAATGGAGGCGGAATTA	CCTCTGACTCCCACAGGAAG	3C assay
B-15	TCATGGTCTAAGGCTGAGCA	TGCTTACCACAGTGGCATAAA	3C assay
ECORI site between HIF binding and CCND1	CTAGGGCATGACCACGATG	GCCAAAGTCTCAGGTGGAAA	3C assay
B-30	GAATGAACGAGTACAGGGGG	TTTTTGTTTCGACTTGATAGCC	3C assay
hu CCND1 promoter ECORI 5prime 1	GAGTAGCCAGGGCAACCATA	CAGTCCTTACCCCAGAAGCA	3C assay
hu CCND1 promoter ECORI 5prime 6	CCCAGGACCCGGAATATTAG	GGGTGCTCACAGCAAGATG	3C assay
hu CCND1 promoter ECORI 5prime 7	GTTCTGAATTATGCCGGCTC	GGTTTTGTTGGGGGTGTAGA	3C assay
hu CCND1 promoter ECORI 5prime 8	CCCTTGCGAGAGGGAAAC	GTTGGGAGTTCGTTACAGGC	3C assay
hu CCND1 promoter ECORI 5prime 9	ACACTCATCCCTGCCGAC	GTGTGGGATGTGCCTCTTG	3C assay
EEF1G	GTTCGCCCACCAGAAAAGTC	GGCCATTAAAATGTTTGCAAAGA	3C assay
EEF1G probe	FAM-CAAGTAAGCATC CAGCAGCCCCAGAA-BHQ-1		3C assay
hu CCND1 HIF2 ECORI 3prime B	AGGCCAACAGAGAAGAGTGG	CTTGCCACATTGCTTTCAGA	3C assay/FAIRE
NDRG1 HIF binding ECORI B	TCCCTAGGGGAAAAGGACAC	TGATTTCTAACCAGGCAGCA	FAIRE

CCND1 HIF bind upstream1	GCTGGGATTATAGGCGTGAG	CCAGCACTGTGCAATAGGAA	FAIRE
CCND1 HIF bind upstream2	TGCGCCATTTTACAGACAAG	TGGAGATTCTGGGACTTTGC	FAIRE
CCND1 HIF bind upstream3	GTGGAGGGCATCTTTCTGAG	CAAGACTGAGGGTTGCCAAT	FAIRE
CCND1 HIF bind 5 prime 2	CCAGATTCAGACCCTTCCAG	TTGACCAGAAACCCGAAAAA	FAIRE
CCND1 HIF bind 5 prime 3	TTTTTCGGGTTTCTGGTCAA	CCTCCCATCAGGCAAAAATA	FAIRE
CCND1 HIF bind 5 prime 5	TTCCGTGGCAACATTACATC	CTCAGTGTCCGTGACTGTGG	FAIRE
CCND1 HIF bind 5 prime 8	CCAAATTGGCAGCACTTTCT	TGCATGTGCATAAACACTCG	FAIRE
CCND1 HIF bind downstream 1	TGCAGATACGTCTGCCTCTC	TCCTCCCTAAAGGCAACTGA	FAIRE
CCND1 HIF bind downstream 2	CTGTGAGGCTGTGATTCTGG	GCCTGACCACAGGACATTTT	FAIRE
CCND1 HIF bind downstream 3	TTCCCCATTCACAGGAGAAG	GCATCACAGAACAGCTGCAT	FAIRE
CCND1 HIF bind downstream 4	ATGCAGCTGTTCTGTGATGC	TTCATGGAGGAGAAACTGGG	FAIRE
CCND1 HIF bind downstream 5	GTATCCGTGCTTGGGATGAT	ATCGTGTCAGTGATGGGACA	FAIRE
CCND1 HIF bind downstream 6	TGTGGTGGATCAACGTCACT	GACACAGATGGCTGGGTTTT	FAIRE
CCND1 HIF bind downstream 8	CACCACGCCCACTTAATTTT	CTTTGGGAGGCTGAGACAAG	FAIRE
CCND1 HIF bind downstream 9	GGAAGGGTTGGTTAGCATGA	GCCCTGGCTCAGTAGTTGTT	FAIRE
ARRDC3 expression	AATTCCGAAGAAGGCTTCCA	CGAATGAGGTAGCGAGTGG	Expression qPCR
ERRFI1 expression	CCGATAACCATGGCCTACAG	ATTCATCGGAGCAGATTTGG	Expression qPCR
HIF1A antisense expression	GCCCCTTTGATCAGCTTTTA	GAGTTGGAGGTGTTGAAGCA	Expression-qPCR
Nici expression 5’ peak	CTTCTTGTTAGGGGATGCCA	CGCCCTAATCTCAGCCTTTT	Expression-qPCR
Nici expression middle peak	CCCTTCGTCTCTGTTTCTCG	CTGAGACCCTGCCTCAAAAG	Expression-qPCR
Nici expression 3’ peak	CTCCCTTTCTTGCACTCGTC	GGTAAATGCGTCCAAGAGGA	Expression-qPCR
HIG2 expression	CACGACCTGCTCCTACAGC	GCAGAGAAACAGAGCTGCCT	Expression-qPCR
HPRT expression	GAC CAG TCA ACA GGG GAC AT	AAC ACT TCG TGG GGT CCT TTT C	Expression-qPCR

7.8.2 Taqman® assays

Target transcript	Company	Cat. number	Comment
CCND1	Applied Biosystems	Hs00765553_m1	
CA9	Applied Biosystems	Hs00154208_m1	
VEGF	Applied Biosystems	Hs00900055_m1	

GLUT1	Applied Biosystems	Hs00892681_m1	
FIH	Applied Biosystems	Hs00215495_m1	
HIF-1α	Applied Biosystems	Hs00153153_m1	
Cyclophillin A	Applied Biosystems	4326316E	endogenous control

7.9 Solutions and buffers

7.9.1 Urea- sodium dodecyl sulphate (SDS) buffer

Name	**Stock**	**Volume**	
Urea in H_2O	8 M	8.4 ml	
Glycerol	100%	1 ml	
SDS	20%	0.5 ml	
DTT	1 M	0.05 ml	
TRIS pH 6.8	1 M	0.1 ml	
PMSF	50 mM	0.1 ml	Add before use

7.9.2 Western blotting running buffer (10x)

Name	**Final**
TRIS	0.25 M
Glycine	1.9 M
SDS	1%

7.9.3 Western blotting transfer buffer (20X)

Name	Final
TRIS	0.2 M
Glycine	2 M

7.9.4 SDS-PAGE resolving gel buffer

Name	Final
TRIS pH 8.8	1.5 M
SDS	0.4%

7.9.5 SDS-PAGE stacking gel buffer

Name	Final
TRIS pH 6.8	1.5 M
SDS	0.4%

7.9.6 Phosphate buffered saline (PBS)

Name	Final
NaCl	0.137 M
KCl	0.0027 M
Phosphate buffer	0.1 M

7.9.7 PBS-Tween (PBS-T)

Name	Final
NaCl	0.137 M
KCl	0.0027 M
Phosphate buffer	0.1 M
Tween 20	0.05%

7.9.8 5% Milk in PBS-T

Name	Weight/volume	Final
Skim milk powder	5 g	5%
PBS-T	Fill to 100 ml	

7.9.9 Cell lysis buffer for DNA extraction

Name	Final
NaCl	100 mM
TRIS pH 8.0	10 mM
EDTA	25 mM
SDS	0.5%
Proteinase K	0.1 mg/ml

7.9.10 ChIP-Lysis buffer

Name	Stock	Volume	Final
SDS	20%	0.05 ml	1%
EDTA	0.5 M	0.02 ml	10 mM
TRIS pH 8.1	1 M	0.05 ml	50 mM
Complete inhibitor (Roche)	25x	0.04 ml	1x
H_2O		0.84 ml	

7.9.11 ChIP-Dilution buffer

Name	Stock	Volume	Final
SDS	20%	0.025 ml	0.01%
EDTA	0.5 M	0.12 ml	1.2 mM
TRIS pH 8.1	1 M	0.835 ml	16.7 mM
Triton X-100	100%	0.55 ml	1.1%
NaCl	5 M	1.67 ml	167 mM
Complete inhibitor (Roche)	1 tablet		
H_2O		46.8ml	

7.9.12 Low salt buffer

Name	Stock	Volume	Final
SDS	20%	0.005 ml	0.1%
Triton X-100	100%	0.01 ml	1%
EDTA	0.5 M	0.004 ml	2 mM
TRIS pH 8.1	1 M	0.02 ml	20 mM
NaCl	5 M	0.03 ml	150 mM
H_2O		0.931 ml	

7.9.13 High salt buffer

Name	Stock	Volume	Final
SDS	20%	0.005 ml	0.1%
Triton X-100	100%	0.01 ml	1%
EDTA	0.5 M	0.004 ml	2 mM
TRIS pH 8.1	1 M	0.02 ml	20 mM
NaCl	5 M	0.1 ml	500 mM
H_2O		0.861 ml	

7.9.14 Lithium chloride buffer

Name	Stock	Volume/weight	Final
LiCl	8 M	0.032 ml	0.25 M
Igepal	100%	0.01 ml	1%

Deoxycholate Na		0.01 g	1%
EDTA	0.5 M	0.002 ml	1 mM
TRIS pH 8.1	1 M	0.01 ml	10 mM
H_2O		0.946 ml	

7.9.15 TRIS-EDTA (TE) buffer

Name	**Stock**	**Volume**	**Final**
TRIS pH 8.0	1 M	0.01 ml	10 mM
EDTA	0.5 M	0.002 ml	1 mM
H_2O		0.988 ml	

7.9.16 TRIS-Borate-EDTA (TBE) buffer (5x)

Name	**Stock**	**Volume/weight**
TRIS base		54 g
Boric acid		27.5 g
EDTA	0.5 M	20 ml
H_2O		to 1 l

7.9.17 Elution buffer

Name	Stock	Volume/weight	Final
$NaHCO_3$		0.0841 g	0.1 M
SDS	20%	0.5 ml	1%
H_2O		9.5 ml	

7.9.18 Lysis buffer for 3C assay

Name	Final
TRIS-HCL pH 7.5	10 mM
NaCl	10 mM
NP-40	0.2%
Complete proteinase inhibitor (Roche)	1X

7.9.19 Glucose-EDTA-TRIS (GET) solution

Name	Stock	Volume
Glucose	20%	23 ml
EDTA	0.1 M	50 ml
TRIS pH 7.0	1 M	13 ml
H_2O		420 ml

7.9.20 NaOH/SDS solution

Name	Stock	Volume
NaOH	5 M	20 ml
SDS	20%	25 ml
H_2O		455 ml

7.9.21 Potassium acetate solution (3 M)

Name	Volume/weight
Potassium acetate	147.21 g
Glacial acetic acid	57.7 ml
H_2O	To 500 ml

7.9.22 Z-buffer for β-galactosidase assay

Name		Final	Comments
$Na_2HPO_4x7H_2O$	16.1 g	0.06 M	
$NaH_2PO_4xH_2O$	5.5 g	0.04 M	
KCl	0.75 g	0.01 M	
$MgSO_4x7H_2O$	0.246 g	0.001 M	
Beta-mercaptoethanol	2.7 ml	0.05 M	Add just before assay
H_2O	To 1 l		

8 Appendix

8.1 Conditions for ChIP-seq experiments

ChIP- experiment	Antibody	Conditions
MCF-7 cells		
HIF-1α	PM14	DMOG 2 mM 16 h
HIF-2α	PM9	DMOG 2 mM 16 h
HIF-1β	NB 100-110	Hypoxia 0.5% 16 h, Normoxia
Preimmune	PM14 preimmune	DMOG 2 mM 16 h, Hypoxia 0.5% 16 h, Normoxia
RNA pol2	Sc-899	Hypoxia 0.5% 16 h, Normoxia
H3K4me1	ab8895	Hypoxia 0.5% 16 h, Normoxia
H3K4me3	9751	Hypoxia 0.5% 16 h, Normoxia
786-O cells		
HIF-2α	PM9	Normoxia
HIF-1β	NB 100-110	Normoxia
Preimmune	PM14 preimmune	Normoxia

8.2 High stringency HIF-binding sites

8.2.1 High stringency HIF-1 binding sites in MCF-7 cells

HIF-1α binding sites	chromosome	start	end	Nearest gene locus	RefSeq number
1	chr7	156986147	156987039	DNAJB6	NM_058246
2	chr8	134456859	134457261	NDRG1	NM_006096
3	chr1	114156539	114157002	RSBN1	NM_018364
4	chr16	29984123	29984574	ALDOA	NM_184041
5	chr15	91153222	91153655	CHD2	NM_001042572
6	chr20	48720371	48720786	C20orf175	NM_080829
7	chr3	5003621	5003966	BHLHB2	NM_003670
8	chr14	61287408	61287886	SNAPC1	NM_003082
9	chr10	6283489	6283830	PFKFB3	NM_004566
10	chr19	39541937	39542408	GPI	NM_000175
11	chr1	178403857	178404013	QSCN6	NM_002826
12	chr15	70306014	70306475	PKM2	NM_002654
13	chr6	87921667	87922197	CGA	NM_000735
14	chr14	74791306	74791781	FOS	NM_005252
15	chr14	22840340	22840855	BCL2L2	NM_004050
16	chr14	33477245	33477551	EGLN3	NM_022073
17	chr6	43743806	43744203	MRPS18A	NM_018135
18	chr12	8014473	8014830	SLC2A3	NM_006931
19	chr3	48569558	48569870	PFKFB4	NM_004567
20	chr7	55216284	55216762	EGFR	NM_201283
21	chr2	173128811	173129189	PDK1	NM_002610

22	chr4	186554370	186554738	ANKRD37	NM_181726
23	chr3	123585527	123585631	C3orf28	NM_014367
24	chr8	135913721	135913971	ZFAT1	NM_020863
25	chr1	8861349	8862115	ENO1	NM_001428
26	chr17	78009247	78009655	NARF	NM_001038618
27	chr1	200320766	200321048	GPR37L1	NM_004767
28	chr2	86521270	86521896	JMJD1A	NM_018433
29	chr15	88509909	88510331	SEMA4B	NM_020210
30	chr4	120441100	120441561	LOC401152	NM_001001701
31	chr7	74146696	74146993	GTF2IRD2B	NM_001003795
32	chr7	127882809	127883132	HIG2	NM_013332
33	chr17	18026927	18027385	ALKBH5	NM_017758
34	chr6	89848086	89848478	PNRC1	NM_006813
35	chr2	136458976	136459346	DARS	NM_001349
36	chr11	68638847	68639275	TPCN2	NM_139075
37	chr22	29149882	29150322	MTP18	NM_016498
38	chr2	218226245	218226679	TNS1	NM_022648
39	chr15	72694839	72695207	CLK3	NM_001292
40	chr14	33599049	33599443	EGLN3	NM_022073
41	chr1	200258705	200259075	ELF3	NM_004433
42	chrX	77245930	77246317	PGK1	NM_000291
43	chr5	172688961	172689320	STC2	NM_003714
44	chr12	123888676	123889064	SCARB1	NM_005505
45	chr12	53112524	53112851	ITGA5	NM_002205
46	chr10	75202572	75202889	FUT11	NM_173540
47	chr8	128952244	128952532	TMEM75	NM_001037234
48	chr3	81892736	81893152	GBE1	NM_000158
49	chr5	90612044	90612260	ARRDC3	NM_020801
50	chr11	62024325	62024606	AHNAK	NM_001620
51	chr11	18372393	18372802	LDHA	NM_005566
52	chr1	201047064	201047357	JARID1B	NM_006618
53	chr3	11658884	11659327	VGLL4	NM_014667
54	chr7	104440272	104440545	MLL5	NM_182931
55	chr8	28325357	28325628	ZNF395	NM_018660
56	chr4	174527365	174527591	SAP30	NM_003864
57	chr11	72583327	72583683	P2RY2	NM_002564
58	chr3	4995801	4996077	BHLHB2	NM_003670
59	chr14	69582494	69582887	SLC8A3	NM_182936
60	chr21	42669540	42669888	TFF1	NM_003225
61	chr14	73292926	73293262	C14orf43	NM_194278
62	chr12	50823773	50823987	KRT80	NM_182507
63	chr16	78618607	78618938	MAF	NM_005360
64	chr3	147361715	147361989	PLOD2	NM_000935
65	chr10	74526498	74526883	P4HA1	NM_000917
66	chr9	129380930	129381235	C9orf88	NM_001035534
67	chr7	73905100	73905444	GTF2IRD2	NM_173537
68	chr14	99628980	99629302	EVL	NM_016337
69	chr20	49800124	49800349	ATP9A	NM_006045
70	chr6	121912548	121912909	GJA1	NM_000165
71	chr14	64763490	64763783	MAX	NM_145114
72	chr11	75833202	75833572	C11orf30	NM_020193
73	chr7	156890448	156890830	DNAJB6	NM_058246
74	chr1	226419772	226420115	C1orf69	NM_001010867
75	chr12	119396854	119397160	DYNLL1	NM_001037494
76	chr15	83259150	83259393	SLC28A1	NM_201651
77	chr3	25806329	25806547	OXSM	NM_017897
78	chr17	23927916	23928095	ALDOC	NM_005165
79	chr16	31013479	31013882	VKORC1	NM_024006
80	chr20	48215897	48216284	Kua	NM_199129
81	chr16	85983157	85983476	MAP1LC3B	NM_022818
82	chr9	6706157	6706472	JMJD2C	NM_015061
83	chr9	83392960	83393210	TLE1	NM_005077

84	chr8	81246341	81246554	TPD52	NM_001025253
85	chr3	158289292	158289559	CCNL1	NM_020307
86	chr8	28279212	28279555	ZNF395	NM_018660
87	chr1	8184878	8185187	ERRFI1	NM_018948
88	chr17	19821692	19821899	AKAP10	NM_007202
89	chr11	64946253	64946440	FKSG44	NM_031904
90	chr17	36945908	36946291	KRT19	NM_002276
91	chr2	201286087	201286398	BZW1	NM_014670
92	chr8	135559830	135560140	ZFAT1	NM_001029939
93	chr12	43913051	43913337	TMEM16F	NM_001025356
94	chr13	24773529	24773764	NUPL1	NM_001008565
95	chr8	22511542	22511842	C8orf58	NM_001013842
96	chr15	99973759	99974053	TM2D3	NM_025141
97	chr10	73690712	73690933	DDIT4	NM_019058
98	chr3	151963508	151963803	SIAH2	NM_005067
99	chr12	6513611	6513887	GAPDH	NM_002046
100	chr11	558467	558680	RASSF7	NM_003475
101	chr16	84236240	84236607	KIAA0182	NM_014615
102	chr17	43381206	43381531	PNPO	NM_018129
103	chr14	33602452	33602755	EGLN3	NM_022073
104	chr16	14365785	14366127	PARN	NM_002582
105	chr16	86381073	86381406	KLHDC4	NM_017566
106	chr2	1614676	1614975	PXDN	NM_012293
107	chr4	124537876	124538123	SPRY1	NM_199327
108	chr17	70960822	70961069	KIAA0195	NM_014738
109	chr8	134298157	134298375	WISP1	NM_080838
110	chr3	64065487	64065587	PSMD6	NM_014814
111	chr5	130999015	130999361	RAPGEF6	NM_016340
112	chr14	68205669	68206009	ZFP36L1	NM_004926
113	chr19	45028745	45029040	FBL	NM_001436
114	chr12	32988438	32988726	PKP2	NM_001005242
115	chr3	144803636	144803858	SLC9A9	NM_173653
116	chr2	118562037	118562257	INSIG2	NM_016133
117	chr19	18529505	18529725	C19orf50	NM_024069
118	chr22	36422654	36422873	TRIOBP	NM_001039141
119	chr11	10285961	10286173	ADM	NM_001124
120	chr10	6381458	6381580	PFKFB3	NM_004566
121	chr2	70223275	70223697	C2orf42	NM_017880
122	chr2	666606	666895	TMEM18	NM_152834
123	chr6	7506718	7506999	DSP	NM_001008844
124	chr8	23258802	23259078	LOXL2	NM_002318
125	chr10	6381207	6381456	PFKFB3	NM_004566
126	chr4	186554782	186555017	ANKRD37	NM_181726
127	chr19	45016272	45016491	DYRK1B	NM_004714
128	chr3	16639402	16639613	DAZL	NM_001351
129	chr14	67232180	67232364	RDH11	NM_016026
130	chr5	90612262	90612437	ARRDC3	NM_020801
131	chrY	10633196	10633355	LOC728137	NM_001077697
132	chr20	45526515	45526654	NCOA3	NM_006534
133	chr6	106350610	106350979	PRDM1	NM_001198
134	chr14	58781040	58781395	DAAM1	NM_014992
135	chr3	81893462	81893742	GBE1	NM_000158
136	chr14	64523464	64523703	FNTB	NM_002028
137	chr9	123360097	123360328	DAB2IP	NM_032552
138	chr9	125732302	125732500	DENND1A	NM_024820
139	chr5	114908557	114908722	FEM1C	NM_020177
140	chr2	39204891	39205178	SOS1	NM_005633
141	chr6	170183745	170183951	C6orf122	NM_207502
142	chr2	74914341	74914547	HK2	NM_000189
143	chr11	46920369	46920712	C11orf49	NM_001003676
144	chr8	134193567	134193863	SLA	NM_001045557
145	chr5	71079928	71080218	CARTPT	NM_004291

146	chr19	15102096	15102337	ILVBL	NM_006844
147	chr10	73703426	73703662	DDIT4	NM_019058
148	chr1	148234592	148234825	OTUD7B	NM_020205
149	chr17	42662166	42662382	MYL4	NM_002476
150	chr7	105712797	105713123	PBEF1	NM_005746
151	chr12	6512813	6513123	GAPDH	NM_002046
152	chr19	13806197	13806504	ZSWIM4	NM_023072
153	chr4	89963859	89964130	FAM13A1	NM_001015045
154	chr17	42934695	42934949	NPEPPS	NM_006310
155	chr6	144923150	144923368	UTRN	NM_007124
156	chr10	99175716	99175916	PGAM1	NM_002629
157	chr10	93383011	93383211	PPP1R3C	NM_005398
158	chr7	41579227	41579404	INHBA	NM_002192
159	chrY	10647838	10647940	LOC728137	NM_001077697
160	chr1	114855369	114855610	TRIM33	NM_015906
161	chr2	68035016	68035248	C1D	NM_006333
162	chr3	13348649	13348859	NUP210	NM_024923
163	chr1	148521353	148521520	C1orf51	NM_144697
164	chr7	54794334	54794491	SEC61G	NM_001012456
165	chr4	68249526	68249680	UBE1L2	NM_018227
166	chr16	70075370	70075491	ZNF19	NM_006961
167	chr2	118562259	118562374	INSIG2	NM_016133
168	chr1	16551424	16551742	FBXO42	NM_018994
169	chr19	2475956	2476237	GADD45B	NM_015675
170	chr7	154589369	154589602	HTR5A	NM_024012
171	chr2	220149947	220150166	INHA	NM_002191
172	chr7	1418923	1419127	MICAL-L2	NM_182924
173	chr22	37978168	37978372	PDGFB	NM_002608
174	chr1	201262226	201262430	C1orf37	NM_138391
175	chr2	48521184	48521382	CCDC128	NM_152994
176	chr3	4433633	4433821	SUMF1	NM_182760
177	chr1	148521522	148521672	C1orf51	NM_144697
178	chr20	48840670	48840807	BCAS4	NM_017843
179	chr3	63983969	63984098	PSMD6	NM_014814
180	chr17	33696104	33696361	MRPL45	NM_032351
181	chr9	35663732	35663974	CA9	NM_001216
182	chr20	33668744	33668974	SPAG4	NM_003116
183	chr6	52476651	52476875	TRAM2	NM_012288
184	chr19	5671091	5671289	LONP1	NM_004793
185	chr2	216193083	216193274	FN1	NM_054034
186	chr20	52241927	52242054	PFDN4	NM_002623
187	chr20	44151869	44152166	NCOA5	NM_020967
188	chr1	243065104	243065398	FAM36A	NM_198076
189	chr8	103735347	103735608	KLF10	NM_001032282
190	chr4	25930198	25930459	RBPSUH	NM_203283
191	chr21	41459111	41459348	BACE2	NM_138992
192	chr8	80874291	80874527	HEY1	NM_012258
193	chr17	7679626	7679852	TMEM88	NM_203411
194	chr10	102096101	102096315	SCD	NM_005063
195	chr19	16544212	16544417	SLC35E1	NM_024881
196	chr18	70747262	70747462	ZNF407	NM_017757
197	chr11	10286194	10286392	ADM	NM_001124
198	chr9	126743076	126743254	GOLGA1	NM_002077
199	chr11	70688149	70688260	DHCR7	NM_001360
200	chr2	132741722	132741829	GPR39	NM_001508
201	chr3	10132123	10132347	C3orf10	NM_018462
202	chr9	115023395	115023610	SLC31A1	NM_001859
203	chr2	134557694	134557905	MGAT5	NM_002410
204	chr10	78934794	78935001	KCNMA1	NM_002247
205	chr6	169920917	169921112	C6orf208	NM_025002
206	chr17	22645082	22645276	WSB1	NM_015626
207	chr7	55221555	55221748	EGFR	NM_201283

208	chr12	9841911	9842054	KLRF1	NM_016523
209	chr9	15412582	15412719	SNAPC3	NM_001039697
210	chr8	26364721	26364847	PNMA2	NM_007257
211	chr7	50856665	50856947	GRB10	NM_001001555
212	chr9	116287020	116287264	DFNB31	NM_015404
213	chr7	156624126	156624369	UBE3C	NM_014671
214	chr6	157384220	157384453	ARID1B	NM_017519
215	chr10	4758298	4758530	AKR1CL2	NM_001040177
216	chr7	101403175	101403373	CUTL1	NM_181552
217	chr14	58898000	58898198	GPR135	NM_022571
218	chr2	219871791	219871988	PTPRN	NM_002846
219	chr4	153449508	153449703	FBXW7	NM_018315
220	chr9	84867979	84868162	RASEF	NM_152573
221	chr19	18352006	18352189	GDF15	NM_004864
222	chr5	179092228	179092410	MAML1	NM_014757
223	chr3	108726344	108726516	BBX	NM_020235
224	chr1	159328997	159329167	PVRL4	NM_030916
225	chr3	197293433	197293600	TFRC	NM_003234
226	chr11	73149848	73150003	RAB6A	NM_002869
227	chr1	150228700	150228850	S100A10	NM_002966
228	chr20	48780773	48780916	PARD6B	NM_032521
229	chr17	58231724	58231944	RNF190	NM_152598
230	chr5	133368521	133368737	VDAC1	NM_003374
231	chr13	76452268	76452464	CLN5	NM_006493
232	chr15	62025022	62025211	DAPK2	NM_014326
233	chr8	134297949	134298128	WISP1	NM_080838
234	chr14	74795772	74795951	FOS	NM_005252
235	chr10	70330915	70331083	DDX50	NM_024045
236	chr2	74502257	74502421	WDR54	NM_032118
237	chr17	56001980	56002130	PPM1D	NM_003620
238	chr2	240402920	240403200	NDUFA10	NM_004544
239	chr10	127515177	127515441	BCCIP	NM_078468
240	chr1	166988060	166988278	DPT	NM_001937
241	chr22	16020109	16020317	CECR5	NM_033070
242	chr14	53474991	53475189	BMP4	NM_130851
243	chr10	104149321	104149519	NFKB2	NM_002502
244	chr21	35029053	35029248	CLIC6	NM_053277
245	chr1	154147887	154148076	RIT1	NM_006912
246	chr4	6721001	6721182	S100P	NM_005980
247	chr6	2933759	2933936	DKFZp686I15217	NM_207495
248	chr20	39402887	39403063	LPIN3	NM_022896
249	chr1	116455850	116456013	C1orf161	NM_152367
250	chr22	41853572	41853727	MCAT	NM_173467
251	chr7	63784350	63784501	ZNF588	NM_001013746
252	chr12	129922294	129922422	RAN	NM_006325
253	chr15	72969438	72969558	MPI	NM_002435
254	chr16	1410753	1410863	UNKL	NM_001037125
255	chr20	48811820	48811930	PARD6B	NM_032521
256	chr1	191357449	191357762	CDC73	NM_024529
257	chr5	1227455	1227752	SLC6A19	NM_001003841
258	chr14	64523232	64523462	FNTB	NM_002028
259	chr17	77412195	77412423	P4HB	NM_000918
260	chr2	28828076	28828299	PPP1CB	NM_002709
261	chr19	52938529	52938721	GLTSCR2	NM_015710
262	chr1	43200941	43201132	SLC2A1	NM_006516
263	chr1	199833538	199833724	NAV1	NM_020443
264	chr3	151545848	151546032	TSC22D2	NM_014779
265	chr19	10720894	10721075	DNM2	NM_001005360
266	chr6	116515250	116515417	NT5DC1	NM_152729
267	chr14	73149012	73149174	ACOT6	NM_001037162
268	chr2	64723977	64724136	SERTAD2	NM_014755
269	chr12	26158620	26158765	BHLHB3	NM_030762

270	chr1	226394460	226394602	GUK1	NM_000858
271	chr7	45984419	45984709	IGFBP3	NM_000598
272	chr1	205260080	205260321	C1orf116	NM_023938
273	chr10	3101137	3101335	PFKP	NM_002627
274	chr8	125935184	125935381	ZNF572	NM_152412
275	chr6	139738525	139738720	CITED2	NM_006079
276	chr16	65769541	65769736	NOL3	NM_003946
277	chr6	112425017	112425205	WISP3	NM_003880
278	chr9	102228639	102228819	C9orf30	NM_080655
279	chr6	44149100	44149279	MRPL14	NM_032111
280	chr11	2890446	2890622	SLC22A18AS	NM_007105
281	chr3	4995424	4995591	BHLHB2	NM_003670
282	chr19	50600124	50600270	PPP1R13L	NM_006663
283	chr19	4919580	4919724	JMJD2B	NM_015015
284	chr12	69325737	69325872	PTPRB	NM_002837
285	chr6	151430617	151430815	AKAP12	NM_005100
286	chr7	143028336	143028526	FLJ40722	NM_173678
287	chr2	37026	37214	FAM110C	NM_001077710
288	chr17	55575409	55575586	CA4	NM_000717
289	chr5	95429516	95429680	ELL2	NM_012081
290	chr20	52245688	52245840	PFDN4	NM_002623
291	chr13	30268367	30268519	ALOX5AP	NM_001629
292	chr20	45521628	45521778	NCOA3	NM_006534
293	chr13	44813408	44813557	TPT1	NM_003295
294	chr15	90989280	90989426	LOC400451	NM_207446
295	chr1	165456007	165456151	POU2F1	NM_002697
296	chr14	67356870	67357013	RAD51L1	NM_133510
297	chr12	55845773	55845915	LRP1	NM_002332
298	chr3	8649297	8649438	C3orf32	NM_015931
299	chr16	82687693	82687828	MBTPS1	NM_201268
300	chr2	218971496	218971625	CTDSP1	NM_021198
301	chr7	115939423	115939547	CAV1	NM_001753
302	chr21	36724407	36724529	CLDN14	NM_012130
303	chr11	66416401	66416513	PC	NM_022172
304	chr7	27283760	27283910	EVX1	NM_001989
305	chr3	63898670	63898798	ATXN7	NM_000333
306	chr1	174931880	174932000	PAPPA2	NM_021936
307	chr5	134290547	134290659	PCBD2	NM_032151
308	chr5	3979197	3979390	IRX1	NM_024337
309	chr8	134293611	134293861	WISP1	NM_080838
310	chr14	47158437	47158633	MAMDC1	NM_182830
311	chr10	17527768	17527962	ST8SIA6	NM_001004470
312	chr12	45173990	45174172	SLC38A2	NM_018976
313	chr3	38181328	38181504	OXSR1	NM_005109
314	chr9	105896443	105896612	SMC2	NM_001042550
315	chr19	53516939	53517108	EMP3	NM_001425
316	chr1	8194779	8194940	ERRFI1	NM_018948
317	chr14	63178031	63178191	WDR89	NM_080666
318	chr15	50098627	50098784	MAPK6	NM_002748
319	chr7	158349425	158349569	WDR60	NM_018051
320	chr3	158017521	158017663	FLJ16641	NM_001004316
321	chr5	34622102	34622241	RAI14	NM_015577
322	chr17	55566620	55566752	CA4	NM_000717
323	chr3	62280587	62280715	C3orf14	NM_020685
324	chr11	65021534	65021658	SCYL1	NM_020680
325	chr3	101694123	101694246	TMEM45A	NM_018004
326	chr20	45638988	45639111	NCOA3	NM_006534
327	chr9	97678022	97678138	LOC375748	NM_001010895
328	chr8	6553588	6553701	AGPAT5	NM_018361
329	chr17	54538885	54538998	TRIM37	NM_015294
330	chr19	52938324	52938435	GLTSCR2	NM_015710
331	chr1	150232871	150232972	S100A10	NM_002966

332	chr1	184611334	184611504	C1orf27		NM_017847
333	chr3	123617461	123617627	WDR5B		NM_019069
334	chr5	177476080	177476245	N4BP3		NM_015111
335	chr21	44544659	44544811	PFKL		NM_002626
336	chr20	35604091	35604225	BLCAP		NM_006698
337	chr17	7962530	7962654	ALOXE3		NM_021628
338	chr14	96716495	96716614	VRK1		NM_003384
339	chr6	30828142	30828257	IER3		NM_003897
340	chr7	35807041	35807154		39326	NM_001011553
341	chr11	63769939	63770052	PPP1R14B		NM_138689
342	chr3	64098306	64098417	PRICKLE2		NM_198859
343	chr5	131591443	131591553	P4HA2		NM_001017974
344	chr11	129203574	129203684	TMEM45B		NM_138788
345	chr5	140460185	140460294	PCDHB3		NM_018937
346	chr3	59223951	59224059	FLJ42117		NM_198463
347	chr17	56011872	56011980	PPM1D		NM_003620
348	chr2	219803370	219803476	ANKZF1		NM_001042410
349	chr17	55599959	55600059	CA4		NM_000717
350	chr7	1510742	1510922	MAFK		NM_002360
351	chr3	158013018	158013167	FLJ16641		NM_001004316
352	chr3	4995264	4995410	BHLHB2		NM_003670
353	chr15	73017394	73017538	COX5A		NM_004255
354	chr4	78298518	78298661	CCNG2		NM_004354
355	chr3	120386460	120386603	UPK1B		NM_006952
356	chr22	22566538	22566679	MIF		NM_002415
357	chr6	97479227	97479361	KIAA1900		NM_052904
358	chr20	52380545	52380673	PFDN4		NM_002623
359	chr3	197293305	197293431	TFRC		NM_003234
360	chr20	48745097	48745219	PARD6B		NM_032521
361	chr16	84105863	84105985	KIAA0182		NM_014615
362	chr19	614189	614305	RNF126		NM_017876
363	chr17	24253923	24254039	DHRS13		NM_144683
364	chr7	139662810	139662921	SLC37A3		NM_032295
365	chr3	64496987	64497098	ADAMTS9		NM_182920
366	chr8	49083461	49083571	UBE2V2		NM_003350
367	chr17	54442087	54442197	TRIM37		NM_015294
368	chr4	49009903	49010008	FLJ21511		NM_025087
369	chr12	123967669	123967773	UBC		NM_021009
370	chr16	3636756	3636859	DNASE1		NM_005223
371	chr8	101391611	101391712	RNF19		NM_183419
372	chr22	19601163	19601263	CRKL		NM_005207
373	chr18	7940851	7940993	PTPRM		NM_002845
374	chr11	47620700	47620840	MTCH2		NM_014342
375	chr13	40008701	40008839	FOXO1A		NM_002015
376	chr2	38177015	38177144	CYP1B1		NM_000104
377	chr17	2665892	2666014	GARNL4		NM_015085
378	chr16	2029618	2029729	NTHL1		NM_002528
379	chr7	138371287	138371397	ZC3HAV1L		NM_080660
380	chr21	35182336	35182443	RUNX1		NM_001001890
381	chr6	36669894	36670000	SFRS3		NM_003017
382	chr3	23745429	23745535	UBE2E1		NM_003341
383	chrX	67976839	67976944	EFNB1		NM_004429
384	chr20	52651541	52651646	DOK5		NM_018431
385	chr19	58715722	58715827	ZNF331		NM_018555
386	chr19	57384625	57384730	PPP2R1A		NM_014225
387	chr4	48979885	48979988	FLJ21511		NM_025087
388	chr11	65341954	65342057	FLJ30934		NM_152760
389	chr7	72735729	72735831	WBSCR22		NM_017528
390	chr20	48731749	48731850	C20orf175		NM_080829
391	chr2	156747043	156747142	NR4A2		NM_173173
392	chr19	39121672	39121771	KCTD15		NM_024076
393	chr13	38726935	38727034	LOC387921		NM_001012754

394	chr12	2792228	2792327	ITFG2	NM_018463
395	chr17	72233583	72233710	PTDSR	NM_015167
396	chr17	55353277	55353379	RPS6KB1	NM_003161
397	chr12	123473700	123473802	NCOR2	NM_006312
398	chr1	115060855	115060956	NRAS	NM_002524
399	chr19	8332927	8333027	ANGPTL4	NM_001039667
400	chr20	48739889	48739988	PARD6B	NM_032521

8.2.2 High stringency HIF-2 binding sites in MCF-7 cells

HIF-2α binding sites MCF-7	chromosome	start	end	Nearest gene locus	RefSeq number
1	chr7	156986122	156987013	DNAJB6	NM_058246
2	chr20	48720413	48720792	C20orf175	NM_080829
3	chr15	88509900	88510359	SEMA4B	NM_020210
4	chr21	42669493	42669912	TFF1	NM_003225
5	chr8	134456898	134457355	NDRG1	NM_006096
6	chr7	127882809	127883250	HIG2	NM_013332
7	chr2	218226277	218226673	TNS1	NM_022648
8	chr3	4995548	4996080	BHLHB2	NM_003670
9	chr3	5003471	5003914	BHLHB2	NM_003670
10	chr12	8014461	8014804	SLC2A3	NM_006931
11	chr14	61287435	61287871	SNAPC1	NM_003082
12	chr7	156890452	156890887	DNAJB6	NM_058246
13	chr11	72583304	72583693	P2RY2	NM_002564
14	chr12	50823667	50824115	KRT80	NM_182507
15	chr14	99628951	99629206	EVL	NM_016337
16	chr16	29984146	29984554	ALDOA	NM_184041
17	chr1	200320701	200321105	GPR37L1	NM_004767
18	chr7	55216313	55216811	EGFR	NM_201283
19	chr8	22511476	22511809	C8orf58	NM_001013842
20	chr2	1614654	1615038	PXDN	NM_012293
21	chr7	41579152	41579497	INHBA	NM_002192
22	chr4	120441169	120441511	LOC401152	NM_001001701
23	chr5	3979121	3979463	IRX1	NM_024337
24	chr20	48215760	48216285	Kua	NM_199129
25	chr19	39541912	39542320	GPI	NM_000175
26	chr1	200258697	200259111	ELF3	NM_004433
27	chr12	123888666	123889062	SCARB1	NM_005505
28	chr14	74791304	74791735	FOS	NM_005252
29	chr12	32988395	32988693	PKP2	NM_001005242
30	chr15	62024988	62025281	DAPK2	NM_014326
31	chr6	170183621	170184019	C6orf122	NM_207502
32	chr5	90612044	90612426	ARRDC3	NM_020801
33	chr3	151545734	151546074	TSC22D2	NM_014779
34	chr1	178403708	178404122	QSCN6	NM_002826
35	chr10	6381179	6381565	PFKFB3	NM_004566
36	chr10	6283501	6283764	PFKFB3	NM_004566
37	chr6	139738206	139738703	CITED2	NM_006079
38	chr15	83259176	83259514	SLC28A1	NM_201651
39	chr14	33477277	33477570	EGLN3	NM_022073
40	chr8	134298139	134298421	WISP1	NM_080838
41	chr10	17527696	17527967	ST8SIA6	NM_001004470
42	chr15	70306137	70306407	PKM2	NM_002654
43	chr16	14365768	14366214	PARN	NM_002582
44	chr3	124179619	124179918	SEMA5B	NM_001031702
45	chr6	26404411	26404645	HIST1H4H	NM_003543
46	chr1	8194651	8195042	ERRFI1	NM_018948
47	chr1	114156509	114156878	RSBN1	NM_018364
48	chr11	18372416	18372724	LDHA	NM_005566
49	chr14	68205671	68205953	ZFP36L1	NM_004926
50	chr22	29149966	29150241	MTP18	NM_016498
51	chr8	23258782	23259127	LOXL2	NM_002318
52	chr20	49800148	49800479	ATP9A	NM_006045
53	chr9	83392919	83393289	TLE1	NM_005077
54	chr17	61702178	61702502	PRKCA	NM_002737
55	chr6	121912616	121912922	GJA1	NM_000165

56	chr19	12767968	12768203	JUNB	NM_002229
57	chr3	64065338	64065728	PSMD6	NM_014814
58	chr5	172688972	172689298	STC2	NM_003714
59	chr19	10720745	10721053	DNM2	NM_001005360
60	chr2	74821400	74821698	SEMA4F	NM_004263
61	chr10	75202601	75202863	FUT11	NM_173540
62	chr8	80874301	80874618	HEY1	NM_012258
63	chr11	64946146	64946451	FKSG44	NM_031904
64	chr8	128941246	128941669	TMEM75	NM_001037234
65	chr17	36945869	36946253	KRT19	NM_002276
66	chr14	33599147	33599459	EGLN3	NM_022073
67	chr5	92440043	92440349	NR2F1	NM_005654
68	chr5	71079891	71080400	CARTPT	NM_004291
69	chr6	106350596	106350914	PRDM1	NM_001198
70	chr5	135421932	135422218	TGFBI	NM_000358
71	chr6	87921774	87922088	CGA	NM_000735
72	chr8	125935164	125935446	ZNF572	NM_152412
73	chr17	70960804	70961066	KIAA0195	NM_014738
74	chr3	4433574	4433934	SUMF1	NM_182760
75	chr9	37944168	37944489	MCART1	NM_033412
76	chr14	73292929	73293244	C14orf43	NM_194278
77	chr8	103735289	103735595	KLF10	NM_001032282
78	chr3	195795474	195795699	TMEM44	NM_001011655
79	chr12	53112514	53112832	ITGA5	NM_002205
80	chr6	7506663	7507000	DSP	NM_001008844
81	chr3	62280647	62280981	C3orf14	NM_020685
82	chr3	81893498	81893727	GBE1	NM_000158
83	chr21	35028936	35029242	CLIC6	NM_053277
84	chr2	68034961	68035195	C1D	NM_006333
85	chr2	173128921	173129144	PDK1	NM_002610
86	chr14	58781000	58781405	DAAM1	NM_014992
87	chr8	125281363	125281685	C8orf78	NM_182525
88	chr8	82268829	82269145	PAG1	NM_018440
89	chr9	92737349	92737571	SYK	NM_003177
90	chr19	2476020	2476230	GADD45B	NM_015675
91	chr14	52231958	52232159	ERO1L	NM_014584
92	chr22	19601009	19601367	CRKL	NM_005207
93	chr22	37978065	37978399	PDGFB	NM_002608
94	chr4	153449556	153449870	FBXW7	NM_018315
95	chr2	28584784	28585093	PLB1	NM_153021
96	chr5	90390157	90390437	ARRDC3	NM_020801
97	chr10	78934777	78935043	KCNMA1	NM_002247
98	chr19	13806257	13806499	ZSWIM4	NM_023072
99	chr1	166987972	166988319	DPT	NM_001937
100	chr1	151784434	151784735	S100A4	NM_002961
101	chr3	134944220	134944512	TF	NM_001063
102	chr1	8184897	8185145	ERRFI1	NM_018948
103	chr8	135913801	135914025	ZFAT1	NM_020863
104	chr8	128952234	128952447	TMEM75	NM_001037234
105	chr6	52476619	52477004	TRAM2	NM_012288
106	chr7	55221488	55221775	EGFR	NM_201283
107	chr17	43381171	43381523	PNPO	NM_018129
108	chr11	70688004	70688308	DHCR7	NM_001360
109	chr11	68638866	68639285	TPCN2	NM_139075
110	chr12	6512798	6513142	GAPDH	NM_002046
111	chr10	74526544	74526819	P4HA1	NM_000917
112	chr1	232815082	232815414	IRF2BP2	NM_001077397
113	chr1	116455677	116456009	C1orf161	NM_152367
114	chr1	145181534	145181807	CHD1L	NM_004284
115	chr21	37827527	37827745	DYRK1A	NM_001396
116	chr5	138998030	138998246	CXXC5	NM_016463
117	chr1	27739449	27739662	WASF2	NM_006990

118	chr6	143190239	143190437	HIVEP2	NM_006734
119	chr9	83392475	83392672	TLE1	NM_005077
120	chr6	164326737	164327056	QKI	NM_206855
121	chr4	101236284	101236547	DDIT4L	NM_145244
122	chr10	9320169	9320417	LOC389936	NM_001013656
123	chr17	55218751	55218992	TMEM49	NM_030938
124	chr17	7195878	7196117	KCTD11	NM_001002914
125	chr3	39169143	39169379	AXUD1	NM_033027
126	chr2	201286193	201286396	BZW1	NM_014670
127	chr8	134193583	134193781	SLA	NM_001045557
128	chr3	80017226	80017424	ROBO1	NM_002941
129	chr1	21437666	21437864	ECE1	NM_001397
130	chr12	63305516	63305853	RASSF3	NM_178169
131	chr8	26364621	26364916	PNMA2	NM_007257
132	chr11	46920424	46920630	C11orf49	NM_001003676
133	chr12	119396895	119397086	DYNLL1	NM_001037494
134	chr3	147361719	147362053	PLOD2	NM_000935
135	chr5	157841927	157842208	EBF1	NM_024007
136	chr3	14469330	14469611	SLC6A6	NM_003043
137	chr3	178560425	178560664	TBL1XR1	NM_024665
138	chr3	158017523	158017730	FLJ16641	NM_001004316
139	chr20	35604139	35604337	BLCAP	NM_006698
140	chr1	152657300	152657498	IL6R	NM_181359
141	chr3	158012933	158013315	FLJ16641	NM_001004316
142	chr9	129380968	129381172	C9orf88	NM_001035534
143	chr2	38176943	38177146	CYP1B1	NM_000104
144	chr4	89963900	89964096	FAM13A1	NM_001015045
145	chr17	42934710	42934903	NPEPPS	NM_006310
146	chr17	67859888	67860051	SOX9	NM_000346
147	chr2	11556196	11556353	E2F6	NM_198256
148	chr2	20641094	20641374	HS1BP3	NM_022460
149	chr3	16639355	16639620	DAZL	NM_001351
150	chr12	123473637	123473887	NCOR2	NM_006312
151	chr3	190468582	190468820	FAM79B	NM_198485
152	chr3	144803662	144803860	SLC9A9	NM_173653
153	chr12	45173962	45174155	SLC38A2	NM_018976
154	chr16	78618647	78618820	MAF	NM_005360
155	chr11	558489	558640	RASSF7	NM_003475
156	chr14	33602468	33602799	EGLN3	NM_022073
157	chr8	28279261	28279472	ZNF395	NM_018660
158	chr1	17362041	17362252	PADI1	NM_013358
159	chr11	72584048	72584255	P2RY2	NM_002564
160	chr21	41459045	41459243	BACE2	NM_138992
161	chr17	57231338	57231536	BRIP1	NM_032043
162	chr7	45984323	45984517	IGFBP3	NM_000598
163	chr6	30827326	30827514	IER3	NM_003897
164	chr21	42076608	42076793	RIPK4	NM_020639
165	chr1	79630261	79630441	IFI44	NM_006417
166	chr11	27149619	27149794	BBOX1	NM_003986
167	chr2	218229459	218229680	TNS1	NM_022648
168	chr16	86381114	86381333	KLHDC4	NM_017566
169	chr7	2695387	2695584	AMZ1	NM_133463
170	chr16	66829225	66829403	RBM35B	NM_024939
171	chr3	191138094	191138271	LEPREL1	NM_018192
172	chr10	73690701	73690878	DDIT4	NM_019058
173	chr2	109453607	109453773	38960	NM_144710
174	chr7	45984618	45984770	IGFBP3	NM_000598
175	chr14	100609253	100609584	FLJ41170	NM_001004332
176	chr18	18940577	18940812	CABLES1	NM_138375
177	chr8	102518515	102518717	GRHL2	NM_024915
178	chr7	130963021	130963207	PODXL	NM_001018111
179	chr11	9737849	9738027	SWAP70	NM_015055

180	chr5	17305595	17305763	BASP1	NM_006317
181	chr13	95689061	95689229	HS6ST3	NM_153456
182	chr6	43743923	43744083	MRPS18A	NM_018135
183	chr4	144699635	144699774	LOC441046	NM_001011539
184	chr1	108802840	108803231	FAM102B	NM_001010883
185	chr20	45204728	45205078	EYA2	NM_172112
186	chr21	26316975	26317275	APP	NM_000484
187	chr21	40348212	40348481	PCP4	NM_006198
188	chr11	1185443	1185643	TOLLIP	NM_019009
189	chr3	123585462	123585660	C3orf28	NM_014367
190	chr19	52938278	52938464	GLTSCR2	NM_015710
191	chr10	127515167	127515349	BCCIP	NM_078468
192	chr7	74146690	74146867	GTF2IRD2B	NM_001003795
193	chr20	32577461	32577629	DYNLRB1	NM_014183
194	chr14	37134853	37135019	FOXA1	NM_004496
195	chr8	28325391	28325544	ZNF395	NM_018660
196	chr2	83074310	83074453	LOC388965	NM_001013648
197	chr14	47158407	47158655	MAMDC1	NM_182830
198	chr6	11843467	11843665	C6orf105	NM_032744
199	chr19	15102123	15102321	ILVBL	NM_006844
200	chr1	205260165	205260363	C1orf116	NM_023938
201	chr7	104440386	104440583	MLL5	NM_182931
202	chr1	27727176	27727372	WASF2	NM_006990
203	chr8	127653167	127653346	FAM84B	NM_174911
204	chr8	101391579	101391752	RNF19	NM_183419
205	chr7	143164310	143164475	KIAA0738	NM_014719
206	chr1	176891895	176892060	RALGPS2	NM_152663
207	chr11	76177438	76177588	TSKU	NM_015516
208	chr12	104133438	104133561	DIP13B	NM_018171
209	chr8	128750144	128750266	MYC	NM_002467
210	chr1	238826632	238826744	GREM2	NM_022469
211	chr17	56109291	56109708	BCAS3	NM_017679
212	chr1	108800894	108801189	FAM102B	NM_001010883
213	chr16	86444857	86445055	SLC7A5	NM_003486
214	chr10	3159108	3159306	PITRM1	NM_014889
215	chr12	123967629	123967825	UBC	NM_021009
216	chr15	94239999	94240191	NR2F2	NM_021005
217	chr9	116287027	116287211	DFNB31	NM_015404
218	chr5	134290548	134290726	PCBD2	NM_032151
219	chr2	181094209	181094367	UBE2E3	NM_182678
220	chr3	154360968	154361125	RAP2B	NM_002886
221	chr12	88382972	88383129	GALNT4	NM_003774
222	chr7	41113753	41113908	INHBA	NM_002192
223	chr16	79309922	79310064	CDYL2	NM_152342
224	chr12	116065373	116065510	TESC	NM_017899
225	chr7	33728713	33728849	BMPER	NM_133468
226	chr16	84106192	84106362	KIAA0182	NM_014615
227	chr2	237398784	237398953	CXCR7	NM_020311
228	chr10	35260306	35260473	PARD3	NM_019619
229	chr11	31392580	31392731	DPH4	NM_181706
230	chr6	169920885	169921032	C6orf208	NM_025002
231	chr3	158289325	158289462	CCNL1	NM_020307
232	chr3	126080382	126080518	ITGB5	NM_002213
233	chr7	140324128	140324262	MRPS33	NM_016071
234	chr17	54430391	54430517	TRIM37	NM_015294
235	chr20	52877953	52878059	DOK5	NM_018431
236	chr17	55396893	55397194	LOC51136	NM_016125
237	chr17	57248047	57248337	BRIP1	NM_032043
238	chr1	108579422	108579656	NBPF4	NM_152488
239	chr1	8861639	8861837	ENO1	NM_001428
240	chr12	74658004	74658187	PHLDA1	NM_007350
241	chr3	63701527	63701708	THOC7	NM_025075

242	chr1	183806672	183806842	HMCN1	NM_031935
243	chr16	84236246	84236408	KIAA0182	NM_014615
244	chr5	85616103	85616262	COX7C	NM_001867
245	chrX	77246046	77246194	PGK1	NM_000291
246	chr20	48490804	48490948	PTPN1	NM_002827
247	chr2	217940077	217940218	TNP1	NM_003284
248	chr14	58898005	58898144	GPR135	NM_022571
249	chr10	9151548	9151670	LOC389936	NM_001013656
250	chr7	23340909	23341015	C7orf30	NM_138446
251	chr12	57100613	57100718	XRCC6BP1	NM_033276
252	chr1	20410843	20410945	UBXD3	NM_152376
253	chr2	67595873	67595974	ETAA1	NM_019002
254	chr9	81294344	81294600	TLE4	NM_007005
255	chr3	14893697	14893892	FGD5	NM_152536
256	chr16	65761706	65761900	NOL3	NM_003946
257	chr5	134287674	134287867	PCBD2	NM_032151
258	chr17	55497540	55497725	ABC1	NM_022070
259	chr9	109045846	109046018	RAD23B	NM_002874
260	chr12	12935456	12935621	GPRC5A	NM_003979
261	chr7	101403222	101403386	CUTL1	NM_181552
262	chr3	62848979	62849143	CADPS	NM_003716
263	chr9	110275563	110275724	ACTL7B	NM_006686
264	chr17	22645056	22645214	WSB1	NM_015626
265	chr3	4994968	4995125	BHLHB2	NM_003670
266	chr5	142100322	142100473	ARHGAP26	NM_015071
267	chr6	56636290	56636435	DST	NM_001723
268	chr9	131284557	131284698	C9orf50	NM_199350
269	chr10	102096132	102096271	SCD	NM_005063
270	chr16	86451804	86451934	SLC7A5	NM_003486
271	chr3	196473550	196473675	C3orf21	NM_152531
272	chr16	87521087	87521208	CBFA2T3	NM_175931
273	chr15	72625703	72625810	ARID3B	NM_006465
274	chr5	177476096	177476202	N4BP3	NM_015111
275	chr1	61454079	61454185	NFIA	NM_005595
276	chr13	69954020	69954122	KLHL1	NM_020866
277	chr3	4995262	4995525	BHLHB2	NM_003670
278	chr10	4695410	4695611	AKR1CL2	NM_001040177
279	chr17	55510267	55510449	ABC1	NM_022070
280	chr4	75806215	75806388	AREG	NM_001657
281	chr1	200254322	200254481	ELF3	NM_004433
282	chr7	143028401	143028555	FLJ40722	NM_173678
283	chr22	36422756	36422908	TRIOBP	NM_001039141
284	chr12	8033961	8034110	FOXJ2	NM_018416
285	chr19	45623391	45623536	SERTAD1	NM_013376
286	chr13	112615915	112616060	MCF2L	NM_024979
287	chr15	85421408	85421542	FLJ32310	NM_152336
288	chr17	7679659	7679792	TMEM88	NM_203411
289	chr11	72582528	72582658	P2RY2	NM_002564
290	chr20	42827849	42827976	KCNK15	NM_022358
291	chr14	37133148	37133269	FOXA1	NM_004496
292	chr3	48569651	48569769	PFKFB4	NM_004567
293	chr3	12461776	12461894	TSEN2	NM_025265
294	chr2	127700868	127700986	FLJ16008	NM_001001665
295	chr2	237319174	237319291	CXCR7	NM_020311
296	chr14	52237243	52237356	ERO1L	NM_014584
297	chr5	90550790	90550902	ARRDC3	NM_020801
298	chr11	72582730	72582842	P2RY2	NM_002564
299	chr2	157788506	157788617	GALNT5	NM_014568
300	chr7	143253916	143254025	OR2F2	NM_001004685
301	chr16	85216133	85216242	FOXL1	NM_005250
302	chr3	8649337	8649445	C3orf32	NM_015931
303	chr17	6209753	6209853	AIPL1	NM_001033054

304	chr3	23745406	23745505	UBE2E1	NM_003341
305	chr10	93382849	93383139	PPP1R3C	NM_005398
306	chr6	159341117	159341310	RSHL2	NM_031924
307	chr17	55842802	55842969	C17orf64	NM_181707
308	chr1	115754902	115755034	NGFB	NM_002506
309	chr5	114908538	114908660	FEM1C	NM_020177
310	chr9	90075870	90075991	SPIN	NM_006717
311	chr3	13348695	13348815	NUP210	NM_024923
312	chr17	59754741	59754859	TEX2	NM_018469
313	chr16	66268150	66268262	LOC388284	NM_001012984
314	chr7	22674129	22674238	IL6	NM_000600
315	chr20	52824196	52824304	DOK5	NM_018431
316	chr4	186554451	186554558	ANKRD37	NM_181726
317	chr13	32598427	32598533	STARD13	NM_178007
318	chr1	223945907	223946013	ENAH	NM_001008493
319	chr21	40632917	40633022	PCP4	NM_006198
320	chr12	6513734	6513839	GAPDH	NM_002046
321	chr11	118288240	118288344	BCL9L	NM_182557
322	chr21	29553249	29553351	BACH1	NM_001186
323	chr12	50908488	50908590	KRT7	NM_005556
324	chr19	52938629	52938730	GLTSCR2	NM_015710
325	chr1	114377352	114377452	OLFML3	NM_020190
326	chr1	95103020	95103120	SLC44A3	NM_152369
327	chr17	55823168	55823267	USP32	NM_032582
328	chr15	97844060	97844159	MEF2A	NM_005587
329	chr15	90989458	90989557	LOC400451	NM_207446
330	chr1	114855303	114855515	TRIM33	NM_015906
331	chr17	54538216	54538414	TRIM37	NM_015294
332	chr12	40043640	40043824	PDZRN4	NM_013377
333	chr17	55575087	55575285	CA4	NM_000717
334	chr1	198386404	198386570	NR5A2	NM_003822
335	chr20	52936690	52936820	DOK5	NM_018431
336	chr17	59752112	59752310	TEX2	NM_018469
337	chr17	48538558	48538786	KIF2B	NM_032559
338	chr1	210707028	210707151	NENF	NM_013349
339	chr17	54431210	54431400	TRIM37	NM_015294
340	chr17	54064431	54064618	TEX14	NM_031272
341	chr17	57162525	57162639	BRIP1	NM_032043
342	chr13	108874557	108874701	IRS2	NM_003749
343	chr7	123200805	123200984	WASL	NM_003941
344	chr12	55807313	55807510	LRP1	NM_002332
345	chr1	196120238	196120348	C1orf53	NM_001024594
346	chr11	62024508	62024696	AHNAK	NM_001620
347	chr5	79982688	79982868	MSH3	NM_002439
348	chr1	203900154	203900266	SLC45A3	NM_033102
349	chr12	124104027	124104224	AACS	NM_023928
350	chr17	55571994	55572118	CA4	NM_000717
351	chr4	190391171	190391278	FRG1	NM_004477
352	chr17	54084232	54084335	TEX14	NM_031272
353	chr17	54124837	54124983	RAD51C	NM_002876
354	chr17	56032252	56032363	PPM1D	NM_003620
355	chr1	142124628	142124862	PPIAL4	NM_178230
356	chr2	11597158	11597278	GREB1	NM_148903
357	chr8	125938191	125938414	ZNF572	NM_152412
358	chr17	58251152	58251301	RNF190	NM_152598
359	chr17	55587222	55587335	CA4	NM_000717
360	chr3	64033682	64033806	PSMD6	NM_014814
361	chr1	236699497	236699694	ZP4	NM_021186
362	chr17	55824367	55824553	USP32	NM_032582
363	chr12	67827267	67827374	CPSF6	NM_007007
364	chr5	79981547	79981776	MSH3	NM_002439
365	chr4	190392701	190392900	FRG1	NM_004477

366	chr10	21719115	21719216	C10orf114	NM_001010911
367	chr2	48636464	48636640	SALF	NM_172311
368	chr12	5068221	5068369	KCNA5	NM_002234
369	chr9	41430748	41430849	C9orf36	NM_015667
370	chr2	138003109	138003293	HNMT	NM_001024074
371	chr5	79982943	79983171	MSH3	NM_002439
372	chr17	55574796	55574980	CA4	NM_000717
373	chr20	40449551	40449677	CHD6	NM_032221
374	chr2	76230070	76230178	C2orf3	NM_003203
375	chr20	45805255	45805360	SULF2	NM_018837
376	chr17	55471350	55471505	ABC1	NM_022070
377	chr6	30828100	30828203	IER3	NM_003897
378	chr1	25303376	25303479	SYF2	NM_015484
379	chr9	102499850	102500046	LOC347273	NM_001018116
380	chr17	56033600	56033725	PPM1D	NM_003620
381	chr9	129824332	129824444	FAM102A	NM_001035254
382	chr8	118078399	118078499	LOC441376	NM_001025357
383	chr1	120834711	120834873	FCGR1B	NM_001004340
384	chr17	21944727	21944934	FAM27L	NM_203392
385	chr16	45367588	45367689	MLCK	NM_182493
386	chr3	97819373	97819554	ARL6	NM_032146
387	chr1	154453145	154453287	PMF1	NM_007221
388	chr7	145325296	145325485	CNTNAP2	NM_014141
389	chr5	93930845	93931027	C5orf36	NM_173665
390	chr3	197293333	197293459	TFRC	NM_003234
391	chr17	56027090	56027202	PPM1D	NM_003620
392	chr17	55847565	55847685	C17orf64	NM_181707
393	chrY	5872916	5873088	TSPY2	NM_022573
394	chr20	39538342	39538488	EMILIN3	NM_052846
395	chr4	190390999	190391110	FRG1	NM_004477
396	chr6	31550490	31550589	HCP5	NM_006674
397	chr3	97818737	97818922	ARL6	NM_032146
398	chr17	57240807	57240954	BRIP1	NM_032043
399	chr5	134291215	134291328	PCBD2	NM_032151
400	chr17	48538272	48538452	KIF2B	NM_032559
401	chr17	56033318	56033430	PPM1D	NM_003620
402	chr11	10487344	10487455	RNF141	NM_016422
403	chr2	168128486	168128588	B3GALT1	NM_020981
404	chrY	11929956	11930131	USP9Y	NM_004654
405	chr8	86903925	86904028	REXO1L1	NM_172239
406	chr11	63527462	63527562	OTUB1	NM_017670
407	chr20	26136928	26137027	C20orf191	NM_001039379
408	chr20	26138163	26138296	C20orf191	NM_001039379
409	chr1	154453040	154453140	PMF1	NM_007221
410	chr4	180876451	180876553	AGA	NM_000027
411	chr1	39253494	39253596	NDUFS5	NM_004552
412	chr4	152146165	152146268	LRBA	NM_006726
413	chr13	64939805	64939904	PCDH9	NM_020403
414	chr5	134290731	134290831	PCBD2	NM_032151
415	chr16	58178200	58178299	GOT2	NM_002080
416	chr20	53592722	53592848	CBLN4	NM_080617
417	chr17	54061245	54061362	TEX14	NM_031272
418	chr20	45724943	45725044	SULF2	NM_018837
419	chr1	115061041	115061141	NRAS	NM_002524
420	chr9	67914016	67914115	ANKRD20A1	NM_032250
421	chr17	44252147	44252246	CALCOCO2	NM_005831
422	chr16	164063	164162	HBA2	NM_000517
423	chr17	72204279	72204380	MXRA7	NM_198530
424	chr9	44010605	44010704	LOC441426	NM_001013727
425	chr1	1605676	1605865	CDC2L2	NM_033534

8.2.3 High stringency HIF-2 binding sites in 786-O cells

HIF-2α binding sites 786-O	chromosome	start	end	Nearest gene locus	RefSeq number
1	chr5	1621462	1622140	AYTL2	NM_024830
2	chr8	123841476	123842061	ZHX2	NM_014943
3	chr9	113843539	113844334	SUSD1	NM_022486
4	chr22	34976273	34976680	APOL1	NM_145343
5	chr8	134297840	134298510	WISP1	NM_080838
6	chr1	8194582	8195154	ERRFI1	NM_018948
7	chr10	73751805	73752300	DNAJB12	NM_001002762
8	chr8	95522368	95522692	RAD54B	NM_012415
9	chr20	45188230	45188671	EYA2	NM_172112
10	chr8	68899581	68899966	CPA6	NM_020361
11	chr11	68943508	68943985	MYEOV	NM_138768
12	chr8	134456838	134457384	NDRG1	NM_006096
13	chr9	120133730	120134136	TLR4	NM_138554
14	chr9	138803109	138803925	TMEM141	NM_032928
15	chr9	85975149	85975561	RMI1	NM_024945
16	chr9	83392887	83393238	TLE1	NM_005077
17	chr6	167291769	167292044	RNASET2	NM_003730
18	chr20	56674085	56674427	STX16	NM_001001433
19	chr2	70677234	70677773	TGFA	NM_003236
20	chr10	14478988	14479505	FRMD4A	NM_018027
21	chr10	78934679	78935065	KCNMA1	NM_002247
22	chr1	15555264	15555762	EFHD2	NM_024329
23	chr19	2038114	2038458	MOBKL2A	NM_130807
24	chr9	120786371	120786711	DBC1	NM_014618
25	chr5	90612021	90612534	ARRDC3	NM_020801
26	chr9	92737318	92737613	SYK	NM_003177
27	chr17	76480369	76480739	CHMP6	NM_024591
28	chr6	12652039	12652443	PHACTR1	NM_030948
29	chr10	44797442	44797935	C10orf10	NM_007021
30	chr9	16237183	16237622	LOC648570	NM_001040068
31	chr12	99550969	99551325	GAS2L3	NM_174942
32	chr15	83259189	83259551	SLC28A1	NM_201651
33	chr5	149978512	149978972	SYNPO	NM_007286
34	chr7	139972587	139972964	DENND2A	NM_015689
35	chr7	156890444	156890726	DNAJB6	NM_058246
36	chr3	191766893	191767324	IL1RAP	NM_134470
37	chr5	10604349	10604756	ROPN1L	NM_031916
38	chr13	99022426	99022791	CLYBL	NM_206808
39	chr3	163909375	163909707	C3orf57	NM_001040100
40	chr16	80077030	80077318	CMIP	NM_030629
41	chr1	166987893	166988331	DPT	NM_001937
42	chr6	11501209	11501620	NEDD9	NM_006403
43	chr22	38687382	38687700	FAM83F	NM_138435
44	chr8	134401859	134402315	NDRG1	NM_006096
45	chr3	96558426	96558783	NSUN3	NM_022072
46	chr5	135421874	135422441	TGFBI	NM_000358
47	chr2	28485690	28486059	FOSL2	NM_005253
48	chr10	64013796	64014311	ZNF365	NM_199452
49	chr22	29288268	29288747	GAL3ST1	NM_004861
50	chr1	43179154	43179626	SLC2A1	NM_006516
51	chr16	77760538	77760994	MAF	NM_005360
52	chr2	123181304	123181647	TSN	NM_004622
53	chr7	2646580	2647036	TTYH3	NM_025250
54	chr1	44059772	44060196	ARTN	NM_057091
55	chr16	80159322	80159633	CMIP	NM_030629

56	chr15	61036411	61036854	TPM1	NM_000366
57	chr4	1065908	1066273	LOC285498	NM_194439
58	chr2	85027069	85027431	KCMF1	NM_020122
59	chr9	19151727	19152084	ADFP	NM_001122
60	chr8	82268777	82269137	PAG1	NM_018440
61	chr14	92701334	92701586	MOAP1	NM_022151
62	chr5	1608413	1608869	AYTL2	NM_024830
63	chr5	149979110	149979562	SYNPO	NM_007286
64	chr2	10388637	10389031	HPCAL1	NM_134421
65	chr16	14365713	14366187	PARN	NM_002582
66	chr7	45984381	45984778	IGFBP3	NM_000598
67	chr12	32354627	32354901	FGD4	NM_139241
68	chr3	178500659	178501212	TBL1XR1	NM_024665
69	chr10	29689384	29689683	LYZL1	NM_032517
70	chr8	55020092	55020389	RGS20	NM_003702
71	chr8	126524986	126525214	TRIB1	NM_025195
72	chr16	29984157	29984502	ALDOA	NM_184041
73	chr14	68163479	68163808	ZFP36L1	NM_004926
74	chr5	106729180	106729627	EFNA5	NM_001962
75	chr9	93597881	93598263	ROR2	NM_004560
76	chr20	53033583	53033905	DOK5	NM_018431
77	chr3	195795351	195795762	TMEM44	NM_001011655
78	chr16	86451741	86452133	SLC7A5	NM_003486
79	chr10	6381222	6381563	PFKFB3	NM_004566
80	chr11	10286167	10286486	ADM	NM_001124
81	chr19	10720636	10721073	DNM2	NM_001005360
82	chr10	73690658	73691009	DDIT4	NM_019058
83	chr9	138811417	138811643	KIAA1984	NM_001039374
84	chr15	88509864	88510332	SEMA4B	NM_020210
85	chr9	126111652	126112072	NEK6	NM_014397
86	chr3	188708360	188708697	RTP4	NM_022147
87	chr3	39169122	39169431	AXUD1	NM_033027
88	chr10	29478932	29479263	LYZL1	NM_032517
89	chr17	10214092	10214372	MYH13	NM_003802
90	chr3	150425638	150426011	CP	NM_000096
91	chr1	8073700	8073946	ERRFI1	NM_018948
92	chr5	142536941	142537429	NR3C1	NM_001020825
93	chr20	53167378	53167719	DOK5	NM_018431
94	chr15	61441498	61441835	CA12	NM_001218
95	chr11	115317355	115317668	IGSF4	NM_014333
96	chr3	173725579	173726032	TNFSF10	NM_003810
97	chr1	8184784	8185151	ERRFI1	NM_018948
98	chr6	166024273	166024581	PDE10A	NM_006661
99	chr6	44086532	44086937	MGC45491	NM_153246
100	chr1	232815029	232815430	IRF2BP2	NM_001077397
101	chr2	136720184	136720583	CXCR4	NM_003467
102	chrX	135662654	135662918	ARHGEF6	NM_004840
103	chr9	126498723	126499003	NR6A1	NM_001489
104	chr5	90613051	90613379	ARRDC3	NM_020801
105	chr15	61451612	61451918	CA12	NM_001218
106	chr9	106870059	106870301	ABCA1	NM_005502
107	chr2	123180709	123181203	TSN	NM_004622
108	chr1	115523453	115523902	TSPAN2	NM_005725
109	chr7	20247505	20247974	7A5	NM_182762
110	chr1	154147871	154148218	RIT1	NM_006912
111	chr19	44158933	44159241	FBXO17	NM_024907
112	chr10	73749657	73750009	DNAJB12	NM_001002762
113	chr15	65223954	65224287	SMAD3	NM_005902
114	chr10	3079269	3079598	PFKP	NM_002627
115	chr2	34755877	34756180	FAM98A	NM_015475
116	chr7	151015633	151015970	PRKAG2	NM_024429

117	chr20	319640	319890	TRIB3	NM_021158
118	chr2	237436441	237436771	COPS8	NM_006710
119	chr2	34756200	34756528	FAM98A	NM_015475
120	chr2	234765027	234765557	SPP2	NM_006944
121	chr1	85845379	85845857	CYR61	NM_001554
122	chr7	115939190	115939645	CAV1	NM_001753
123	chr2	134292184	134292621	FLJ34870	NM_207481
124	chr15	70306091	70306407	PKM2	NM_002654
125	chr3	173340933	173341240	FNDC3B	NM_022763
126	chr2	74821336	74821639	SEMA4F	NM_004263
127	chr13	111734453	111734737	SOX1	NM_005986
128	chr9	126088445	126088854	NEK6	NM_014397
129	chr10	3499605	3499929	PITRM1	NM_014889
130	chr3	4995754	4996059	BHLHB2	NM_003670
131	chr12	121247919	121248171	B3GNT4	NM_030765
132	chr12	44781556	44781904	SFRS2IP	NM_004719
133	chr11	558439	558782	RASSF7	NM_003475
134	chr2	177185749	177186007	MTX2	NM_006554
135	chr2	72285184	72285411	CYP26B1	NM_019885
136	chr1	228769291	228769514	COG2	NM_007357
137	chr6	43743723	43744102	MRPS18A	NM_018135
138	chr8	70718766	70719115	SULF1	NM_015170
139	chr19	2442366	2442695	GADD45B	NM_015675
140	chr9	92995786	92996106	AUH	NM_001698
141	chr3	5003591	5003833	BHLHB2	NM_003670
142	chr5	1011674	1012340	NKD2	NM_033120
143	chr12	91471362	91471707	CLLU1OS	NM_001025232
144	chr17	9252480	9252808	STX8	NM_004853
145	chr10	79250142	79250521	DLG5	NM_004747
146	chr19	4919441	4919782	JMJD2B	NM_015015
147	chr5	159591362	159591635	FABP6	NM_001445
148	chr2	234273558	234273798	UGT1A6	NM_001072
149	chr6	11729312	11729651	C6orf105	NM_032744
150	chr1	203741691	203741993	PCTK3	NM_002596
151	chr1	210725850	210726075	NENF	NM_013349
152	chr8	134438403	134438613	NDRG1	NM_006096
153	chr7	134506768	134506973	C7orf49	NM_024033
154	chr2	207967740	207968117	CREB1	NM_004379
155	chr2	54710533	54710859	SPTBN1	NM_178313
156	chr17	33696079	33696383	MRPL45	NM_032351
157	chr3	156086878	156087224	MME	NM_000902
158	chr17	35722093	35722411	RARA	NM_000964
159	chr4	22580025	22580323	GBA3	NM_020973
160	chr10	17068464	17068837	CUBN	NM_001081
161	chr1	97622435	97622691	DPYD	NM_000110
162	chr10	73726939	73727553	DDIT4	NM_019058
163	chr5	316417	316782	PDCD6	NM_013232
164	chr17	40405671	40406024	C1QL1	NM_006688
165	chr19	46565659	46565908	MGC4093	NM_030578
166	chr7	100647901	100648139	PLOD3	NM_001084
167	chr2	145133700	145133936	ZFHX1B	NM_014795
168	chr1	231561741	231562151	KIAA1804	NM_032435
169	chr1	41737058	41737448	EDN2	NM_001956
170	chr9	138592257	138592589	NOTCH1	NM_017617
171	chr3	54765542	54765870	LRTM1	NM_020678
172	chr5	172814764	172815072	STC2	NM_003714
173	chr9	123360060	123360366	DAB2IP	NM_032552
174	chr17	42934682	42934982	NPEPPS	NM_006310
175	chr5	172228119	172228405	ERGIC1	NM_001031711
176	chr1	94965989	94966364	SLC44A3	NM_152369
177	chr20	48726000	48726329	C20orf175	NM_080829

178	chr1	27741517	27741794	WASF2	NM_006990
179	chr22	28922049	28922417	LIF	NM_002309
180	chr3	124179557	124179905	SEMA5B	NM_001031702
181	chr6	107285625	107285943	QRSL1	NM_018292
182	chr10	97266497	97266738	SORBS1	NM_015385
183	chr20	33668775	33668992	SPAG4	NM_003116
184	chr22	17369840	17370282	PRODH	NM_016335
185	chr7	127882865	127883136	HIG2	NM_013332
186	chr10	14864710	14864932	FAM107B	NM_031453
187	chr3	106603904	106604358	ALCAM	NM_001627
188	chr4	129527663	129528054	PGRMC2	NM_006320
189	chr20	48215873	48216149	Kua	NM_199129
190	chr17	61702226	61702500	PRKCA	NM_002737
191	chrX	35371017	35371235	CXorf22	NM_152632
192	chr5	149864603	149864817	NDST1	NM_001543
193	chr5	149833918	149834203	RPS14	NM_001025070
194	chr2	171439173	171439457	GORASP2	NM_015530
195	chr21	46063341	46063546	PCBP3	NM_020528
196	chr22	48305476	48305863	FLJ44385	NM_207478
197	chr3	4433604	4433884	SUMF1	NM_182760
198	chr14	79065128	79065380	NRXN3	NM_138970
199	chr3	125994639	125994864	UMPS	NM_000373
200	chr7	116133935	116134138	MET	NM_000245
201	chr7	55734127	55734506	FKBP9L	NM_182827
202	chr16	11610832	11611152	LITAF	NM_004862
203	chr9	89177238	89177532	DAPK1	NM_004938
204	chr17	72763120	72763405	Sep-09	NM_006640
205	chr19	57930472	57930745	ZNF611	NM_030972
206	chr16	69306605	69306877	LOC92154	NM_138383
207	chr9	92999380	92999636	AUH	NM_001698
208	chr21	39724315	39724555	C21orf13	NM_152505
209	chr22	35054587	35055024	MYH9	NM_002473
210	chr8	131345575	131345962	DDEF1	NM_018482
211	chr10	25391361	25391694	THNSL1	NM_024838
212	chr21	44492285	44492554	ICOSLG	NM_015259
213	chr2	74945958	74946318	HK2	NM_000189
214	chr12	119396796	119397145	DYNLL1	NM_001037494
215	chr5	115907168	115907512	SEMA6A	NM_020796
216	chr2	28584646	28584979	PLB1	NM_153021
217	chr5	131629209	131629500	PDLIM4	NM_003687
218	chr17	58327202	58327480	RNF190	NM_152598
219	chr10	36124112	36124389	FZD8	NM_031866
220	chr1	15176359	15176627	KIAA1026	NM_001017999
221	chr8	142881551	142881771	FLJ43860	NM_207414
222	chr2	227264015	227264407	IRS1	NM_005544
223	chr14	93249174	93249544	PRIMA1	NM_178013
224	chr14	33477262	33477589	EGLN3	NM_022073
225	chr3	85721071	85721397	IGSF4D	NM_153184
226	chr14	33574578	33574856	EGLN3	NM_022073
227	chr16	83234254	83234504	C16orf44	NM_024731
228	chr9	92995051	92995286	AUH	NM_001698
229	chr6	167537441	167537836	TCP10L2	NM_001013683
230	chr2	69814787	69815160	ANXA4	NM_001153
231	chr3	185062998	185063332	PARL	NM_001037639
232	chr8	28279225	28279536	ZNF395	NM_018660
233	chr6	112896238	112896517	LOC442247	NM_001013734
234	chr3	59570754	59570985	FLJ42117	NM_198463
235	chr5	172218383	172218799	ERGIC1	NM_001031711
236	chr1	116464918	116465332	C1orf161	NM_152367
237	chr7	100547376	100547734	SERPINE1	NM_000602
238	chr5	138890995	138891307	UBE2D2	NM_003339

239	chr9	131545298	131545590	PTGES	NM_004878
240	chr11	1808479	1808753	SYT8	NM_138567
241	chr9	93225374	93225630	NFIL3	NM_005384
242	chr2	224389229	224389476	AP1S3	NM_001039569
243	chr20	56631293	56631619	STX16	NM_001001433
244	chr9	92410295	92410601	DIRAS2	NM_017594
245	chr8	141216616	141216828	NIBP	NM_031466
246	chr8	142881799	142881999	FLJ43860	NM_207414
247	chr10	31033543	31033939	LYZL2	NM_183058
248	chr8	23269508	23269827	LOXL2	NM_002318
249	chr20	19809229	19809496	RIN2	NM_018993
250	chr5	116127692	116127917	SEMA6A	NM_020796
251	chr6	11538205	11538577	NEDD9	NM_006403
252	chr8	48734824	48735190	CEBPD	NM_005195
253	chr8	123838496	123838804	ZHX2	NM_014943
254	chr11	57427837	57428143	OR9Q1	NM_001005212
255	chr5	9674703	9674986	TAS2R1	NM_019599
256	chr7	46563544	46563822	IGFBP3	NM_000598
257	chr7	151040037	151040308	PRKAG2	NM_024429
258	chr22	37978075	37978336	PDGFB	NM_002608
259	chr1	67453178	67453437	IL23R	NM_144701
260	chr10	126415285	126415530	FAM53B	NM_014661
261	chr5	55610066	55610289	ANKRD55	NM_024669
262	chr15	99974263	99974470	TM2D3	NM_025141
263	chr9	126086353	126086846	NEK6	NM_014397
264	chr11	125538200	125538614	RPUSD4	NM_032795
265	chr22	25276437	25276750	TPST2	NM_001008566
266	chr8	143778227	143778519	LY6K	NM_017527
267	chr10	75202582	75202862	FUT11	NM_173540
268	chrX	144770880	144771098	CXorf1	NM_004709
269	chr3	14468859	14469060	SLC6A6	NM_003043
270	chr6	123180702	123180901	SMPDL3A	NM_006714
271	chr2	235597590	235597960	SH3BP4	NM_014521
272	chr6	90035219	90035572	GABRR2	NM_002043
273	chr3	135535032	135535352	AMOTL2	NM_016201
274	chr6	167461557	167461850	CCR6	NM_004367
275	chr1	59917059	59917347	HOOK1	NM_015888
276	chr7	149048653	149048927	KRBA1	NM_032534
277	chr1	65386788	65387045	AK3L2	NM_001002921
278	chr10	85921686	85921929	C10orf99	NM_207373
279	chr19	4328270	4328511	SH3GL1	NM_003025
280	chr10	68265857	68266079	LRRTM3	NM_178011
281	chr16	83610126	83610493	ZDHHC7	NM_017740
282	chr10	14550107	14550434	FRMD4A	NM_018027
283	chr2	234951019	234951298	ARL4C	NM_005737
284	chr12	28188262	28188519	CCDC91	NM_018318
285	chr5	167661720	167661947	WWC1	NM_015238
286	chr8	28314473	28314677	ZNF395	NM_018660
287	chr9	113763055	113763433	UGCG	NM_003358
288	chr8	143472914	143473264	TSNARE1	NM_145003
289	chr7	41579173	41579477	INHBA	NM_002192
290	chr1	44041871	44042172	ST3GAL3	NM_174965
291	chr17	53307662	53307944	MRPS23	NM_016070
292	chr8	127202491	127202766	FAM84B	NM_174911
293	chr20	33314769	33314996	ITGB4BP	NM_181468
294	chr19	52165956	52166178	NPAS1	NM_002517
295	chr3	130858291	130858497	TMCC1	NM_015008
296	chr11	34083695	34083893	NAT10	NM_024662
297	chr9	106709307	106709477	ABCA1	NM_005502
298	chr7	55221286	55221742	EGFR	NM_201283
299	chr8	96883244	96883546	GDF6	NM_001001557

300	chr3	147361686	147361980	PLOD2	NM_000935
301	chr5	123229367	123229637	CSNK1G3	NM_004384
302	chr4	2908645	2908901	TETRAN	NM_001120
303	chr16	29938153	29938385	DOC2A	NM_003586
304	chr10	6361479	6361704	PFKFB3	NM_004566
305	chr8	23258885	23259108	LOXL2	NM_002318
306	chr10	95187068	95187289	FER1L3	NM_013451
307	chr22	45506150	45506368	CERK	NM_022766
308	chr3	100860593	100860791	COL8A1	NM_001850
309	chr1	25768250	25768448	LDLRAP1	NM_015627
310	chr21	43845127	43845305	HSF2BP	NM_007031
311	chr5	172214008	172214307	ERGIC1	NM_001031711
312	chr12	50798799	50799083	C12orf44	NM_021934
313	chr9	672885	673110	ANKRD15	NM_153186
314	chr6	45773462	45773660	RUNX2	NM_004348
315	chr15	65155865	65156063	SMAD3	NM_005902
316	chr10	43823809	43824007	ZNF32	NM_006973
317	chr19	6552652	6553001	CD70	NM_001252
318	chr15	99971199	99971505	TM2D3	NM_025141
319	chr12	123888701	123888962	SCARB1	NM_005505
320	chr5	4855307	4855510	ADAMTS16	NM_139056
321	chr12	57557352	57557681	LRIG3	NM_153377
322	chr5	14459664	14459992	FAM105A	NM_019018
323	chr1	12460820	12461147	DHRS3	NM_004753
324	chr5	157794664	157794951	CLINT1	NM_014666
325	chr5	124025055	124025330	ZNF608	NM_020747
326	chr18	21571693	21571956	SS18	NM_001007559
327	chr16	27389630	27389892	IL21R	NM_021798
328	chr10	74526583	74526831	P4HA1	NM_000917
329	chr16	34836702	34836930	LOC649159	NM_001040069
330	chr8	131345272	131345499	DDEF1	NM_018482
331	chr10	7623000	7623216	ITIH5	NM_032817
332	chr3	171927107	171927311	SLC7A14	NM_020949
333	chr8	53770480	53770678	RB1CC1	NM_014781
334	chr6	87921879	87922057	CGA	NM_000735
335	chr5	14089026	14089387	DNAH5	NM_001369
336	chr11	18372425	18372765	LDHA	NM_005566
337	chr22	28912506	28912836	LIF	NM_002309
338	chr10	3139531	3139832	PFKP	NM_002627
339	chr14	99070966	99071260	CCNK	NM_003858
340	chr16	87364420	87364698	FLJ40448	NM_182605
341	chr1	198729052	198729252	ZNF281	NM_012482
342	chr2	88138054	88138482	KRCC1	NM_016618
343	chr7	158349433	158349843	WDR60	NM_018051
344	chr1	17147457	17147809	CROCC	NM_014675
345	chr16	4643178	4643492	MGRN1	NM_015246
346	chr14	47862552	47862856	MAMDC1	NM_182830
347	chr7	48091163	48091449	UPP1	NM_181597
348	chr10	29809263	29809545	SVIL	NM_021738
349	chr9	124186065	124186335	PTGS1	NM_000962
350	chr8	22963370	22963639	TNFRSF10B	NM_003842
351	chr5	10566712	10566968	ROPN1L	NM_031916
352	chr13	112976637	112976862	LAMP1	NM_005561
353	chr2	122174080	122174300	MKI67IP	NM_032390
354	chr1	46570299	46570616	NSUN4	NM_199044
355	chr21	45610160	45610466	COL18A1	NM_130445
356	chr3	45139585	45139878	CDCP1	NM_178181
357	chr10	31024881	31025167	LYZL2	NM_183058
358	chr2	12375997	12376245	TRIB2	NM_021643
359	chr7	139018129	139018608	HIPK2	NM_022740
360	chr21	43972728	43973052	PDXK	NM_003681

361	chr9	139425923	139426184	FLJ20433	NM_017820
362	chr13	43615297	43615541	C13orf21	NM_001010897
363	chr5	125292164	125292396	GRAMD3	NM_023927
364	chr2	201691389	201691587	CFLAR	NM_003879
365	chr2	8631982	8632180	ID2	NM_002166
366	chr12	119167344	119167542	PXN	NM_002859
367	chr5	142605658	142605989	NR3C1	NM_001020825
368	chr1	21437503	21437789	ECE1	NM_001397
369	chr4	79061941	79062220	MRPL1	NM_020236
370	chr10	14478415	14478637	FRMD4A	NM_018027
371	chr1	206045829	206046037	LOC148696	NM_001039568
372	chr1	158146862	158147059	CCDC19	NM_012337
373	chr2	88137305	88137633	KRCC1	NM_016618
374	chr12	69325650	69325940	PTPRB	NM_002837
375	chr4	126870278	126870564	FAT4	NM_024582
376	chr1	8180542	8180827	ERRFI1	NM_018948
377	chr9	19152919	19153201	ADFP	NM_001122
378	chr20	60611216	60611449	FLJ30313	NM_152757
379	chr22	36158203	36158401	MFNG	NM_002405
380	chr15	99973806	99974004	TM2D3	NM_025141
381	chr19	52158261	52158456	GRLF1	NM_004491
382	chr10	3894218	3894404	KLF6	NM_001300
383	chr17	78009228	78009410	NARF	NM_001038618
384	chr1	44072861	44072968	ARTN	NM_057091
385	chr2	218370615	218370923	TNS1	NM_022648
386	chr16	80123750	80124037	CMIP	NM_030629
387	chr4	111625426	111625656	ENPEP	NM_001977
388	chr6	82763941	82764164	FAM46A	NM_017633
389	chr5	168451105	168451314	SLIT3	NM_003062
390	chr2	64724033	64724231	SERTAD2	NM_014755
391	chr18	18940631	18940820	CABLES1	NM_138375
392	chr8	107168477	107168626	OXR1	NM_181354
393	chr6	44148998	44149306	MRPL14	NM_032111
394	chr2	85822479	85822780	ATOH8	NM_032827
395	chr1	28495736	28496006	MED18	NM_017638
396	chr9	37944232	37944456	MCART1	NM_033412
397	chr17	21944728	21944942	FAM27L	NM_203392
398	chr7	75890606	75890816	SRCRB4D	NM_080744
399	chr9	92996143	92996344	AUH	NM_001698
400	chr10	3497863	3498063	PITRM1	NM_014889
401	chr8	103888860	103889058	FLJ45248	NM_207505
402	chr19	4330518	4330716	SH3GL1	NM_003025
403	chr7	128259477	128259672	FLNC	NM_001458
404	chr10	661453	661640	C10orf108	NM_001012714
405	chr8	104419448	104419616	CTHRC1	NM_138455
406	chr11	101852783	101853098	TMEM123	NM_052932
407	chr3	196381382	196381658	C3orf21	NM_152531
408	chr2	86521272	86521537	JMJD1A	NM_018433
409	chr10	73757700	73757925	DNAJB12	NM_001002762
410	chr21	39927609	39927829	C21orf88	NM_153754
411	chr21	33498376	33498582	IFNAR2	NM_000874
412	chr8	22511528	22511733	C8orf58	NM_001013842
413	chr1	198550216	198550420	ZNF281	NM_012482
414	chr5	14318739	14318937	TRIO	NM_007118
415	chr3	45570825	45571023	LIMD1	NM_014240
416	chr2	110229191	110229389	MALL	NM_005434
417	chr6	44085802	44085998	MGC45491	NM_153246
418	chr14	33602484	33602671	EGLN3	NM_022073
419	chr17	42662079	42662264	MYL4	NM_002476
420	chr17	16330800	16330966	C17orf76	NM_207387
421	chr2	235554739	235554884	SH3BP4	NM_014521

422	chr3	36691384	36691724	STAC	NM_003149
423	chr17	4347510	4347847	MYBBP1A	NM_014520
424	chr1	21841094	21841397	RAP1GAP	NM_002885
425	chr16	941771	942071	TMEM112	NM_022773
426	chr20	29612307	29612564	ID1	NM_181353
427	chr2	122781975	122782231	TSN	NM_004622
428	chr16	80302647	80302868	PLCG2	NM_002661
429	chr17	71954643	71954843	UBE2O	NM_022066
430	chr21	35182317	35182514	RUNX1	NM_001001890
431	chr5	1010582	1010726	NKD2	NM_033120
432	chr7	143619469	143619791	OR2A1	NM_001005287
433	chr6	167653401	167653650	TTLL2	NM_031949
434	chr8	82101284	82101526	PAG1	NM_018440
435	chr9	34619472	34619670	ARID3C	NM_001017363
436	chr5	41563044	41563242	PLCXD3	NM_001005473
437	chr7	101492121	101492318	SH2B2	NM_020979
438	chr15	73746201	73746394	SH3PX3	NM_153271
439	chr8	23767527	23767711	STC1	NM_003155
440	chr1	109972311	109972471	AMPD2	NM_203404
441	chr21	15251070	15251220	NRIP1	NM_003489
442	chr10	64421837	64422121	NRBF2	NM_030759
443	chr12	8014543	8014790	SLC2A3	NM_006931
444	chr16	80140418	80140653	CMIP	NM_030629
445	chr3	198605998	198606212	DLG1	NM_004087
446	chrY	17452775	17452980	XKRY2	NM_001002906
447	chr2	234719954	234720152	SPP2	NM_006944
448	chr2	160932841	160933039	RBMS1	NM_016839
449	chr1	44073379	44073577	ARTN	NM_057091
450	chr3	158016582	158016761	FLJ16641	NM_001004316
451	chr2	41937644	41937818	LOC91461	NM_138370
452	chrX	135124318	135124487	FHL1	NM_001449
453	chr6	44066659	44066822	MGC45491	NM_153246
454	chr1	107993238	107993396	VAV3	NM_006113
455	chr15	51910555	51910702	WDR72	NM_182758
456	chr16	83431074	83431437	CRISPLD2	NM_031476
457	chr3	125993175	125993527	UMPS	NM_000373
458	chr11	13118959	13119301	ARNTL	NM_001030272
459	chr12	123886557	123886828	SCARB1	NM_005505
460	chr1	8460081	8460333	RERE	NM_001042682
461	chr7	43493532	43493781	STK17A	NM_004760
462	chr13	20465967	20466191	LATS2	NM_014572
463	chr16	275402	275616	PDIA2	NM_006849
464	chr8	143314343	143314541	TSNARE1	NM_145003
465	chr3	158017455	158017653	FLJ16641	NM_001004316
466	chr2	20658911	20659109	HS1BP3	NM_022460
467	chr16	78075760	78075958	MAF	NM_005360
468	chr10	14487046	14487240	FRMD4A	NM_018027
469	chr7	139091085	139091274	TBXAS1	NM_030984
470	chrX	18775861	18776046	PHKA2	NM_000292
471	chr11	37483238	37483423	RAG2	NM_000536
472	chr12	123955249	123955426	UBC	NM_021009
473	chr7	151136970	151137141	PRKAG2	NM_001040633
474	chr4	72315735	72315896	SLC4A4	NM_003759
475	chr7	4754352	4754628	KIAA0415	NM_014855
476	chr17	74736279	74736555	FLJ21865	NM_001042573
477	chr6	140343503	140343718	CITED2	NM_006079
478	chr1	59121579	59121781	JUN	NM_002228
479	chr3	14468660	14468845	SLC6A6	NM_003043
480	chr8	110031006	110031167	TRHR	NM_003301
481	chr10	101656235	101656396	DNMBP	NM_015221
482	chr22	23734734	23734875	bA9F11.1	NM_001001663

483	chr17	33247502	33247774	TCF2	NM_006481
484	chr12	32262011	32262280	BICD1	NM_001003398
485	chr2	122712056	122712256	TSN	NM_004622
486	chr11	9737924	9738122	SWAP70	NM_015055
487	chr5	169931142	169931325	KCNIP1	NM_001034837
488	chr1	233159838	233160014	TOMM20	NM_014765
489	chr2	228391824	228391997	CCL20	NM_004591
490	chr4	54068854	54069002	LNX1	NM_032622
491	chr20	36245837	36245983	TGM2	NM_198951
492	chr3	197293322	197293436	TFRC	NM_003234
493	chr1	41724627	41724741	EDN2	NM_001956
494	chr9	106873281	106873667	ABCA1	NM_005502
495	chr7	17388734	17389042	AHR	NM_001621
496	chr5	150141556	150141830	MST150	NM_032947
497	chr20	36246018	36246266	TGM2	NM_198951
498	chr6	31262113	31262311	HCG27	NM_181717
499	chr1	19251604	19251802	ALDH4A1	NM_003748
500	chr7	128529873	128530066	TSPAN33	NM_178562
501	chr6	26404237	26404419	HIST1H4H	NM_003543
502	chr3	173792677	173792854	TNFSF10	NM_003810
503	chr2	151050843	151051018	RND3	NM_005168
504	chr20	49800194	49800542	ATP9A	NM_006045
505	chr5	159793554	159793844	PTTG1	NM_004219
506	chr11	697720	697982	EPS8L2	NM_022772
507	chr16	86831165	86831422	ZFPM1	NM_153813
508	chr16	84105911	84106109	KIAA0182	NM_014615
509	chr14	73369580	73369778	ZADH1	NM_152444
510	chr13	29061499	29061696	SLC7A1	NM_003045
511	chr8	17707723	17707906	MTUS1	NM_001001924
512	chr1	59385158	59385336	FLJ10986	NM_018291
513	chr2	237801035	237801208	COPS8	NM_006710
514	chr5	148325735	148325860	SH3TC2	NM_024577
515	chr20	11636994	11637113	BTBD3	NM_181443
516	chr8	131275322	131275428	DDEF1	NM_018482
517	chr16	29185370	29185609	GIYD1	NM_001014999
518	chr2	234865394	234865592	ARL4C	NM_005737
519	chr16	83442378	83442573	CRISPLD2	NM_031476
520	chr10	97378617	97378809	ALDH18A1	NM_001017423
521	chr16	71823903	71824094	C16orf47	NM_207385
522	chr18	3614279	3614465	FLJ35776	NM_001039796
523	chr4	129669834	129669997	PGRMC2	NM_006320
524	chr17	44070296	44070455	HOXB9	NM_024017
525	chr21	41916197	41916353	C21orf129	NM_152506
526	chr16	29938400	29938550	DOC2A	NM_003586
527	chr3	179013871	179014018	TBL1XR1	NM_024665
528	chr10	97267367	97267502	SORBS1	NM_015385
529	chr11	127897830	127897961	ETS1	NM_005238
530	chr10	124031892	124032013	BTBD16	NM_144587
531	chr15	94614919	94615023	NR2F2	NM_021005
532	chr14	89812072	89812284	PSMC1	NM_002802
533	chr2	95374732	95374928	KCNIP3	NM_001034914
534	chr15	72694936	72695127	CLK3	NM_001292
535	chr4	41016414	41016602	DKFZP686A01247	NM_014988
536	chr1	8326258	8326422	RERE	NM_001042682
537	chr16	77763382	77763525	MAF	NM_005360
538	chr8	28265475	28265614	ZNF395	NM_018660
539	chr17	39157027	39157160	SOST	NM_025237
540	chr5	316786	316916	PDCD6	NM_013232
541	chr3	125464770	125464889	KALRN	NM_003947
542	chr3	54765388	54765506	LRTM1	NM_020678
543	chr9	113761943	113762060	UGCG	NM_003358

544	chr9	126088161	126088269	NEK6	NM_014397
545	chr5	1641547	1641655	AYTL2	NM_024830
546	chr3	55169589	55169696	LRTM1	NM_020678
547	chr10	35083895	35084000	PARD3	NM_019619
548	chr2	28469054	28469154	FOSL2	NM_005253
549	chr12	45951405	45951580	FAM113B	NM_138371
550	chr1	230968530	230968690	C1orf57	NM_032324
551	chr11	85107787	85107946	SYTL2	NM_206929
552	chr1	43179696	43179854	SLC2A1	NM_006516
553	chr1	31690242	31690400	SERINC2	NM_178865
554	chr18	44778740	44778885	SMAD7	NM_005904
555	chr1	167351816	167351959	ATP1B1	NM_001677
556	chr8	134510128	134510241	NDRG1	NM_006096
557	chr2	1614719	1614828	PXDN	NM_012293
558	chr4	1313937	1314039	KIAA1530	NM_020894
559	chr17	72763678	72763876	Sep-09	NM_006640
560	chr18	43633562	43633755	SMAD2	NM_005901
561	chr13	36261399	36261573	RFXAP	NM_000538
562	chr2	8631580	8631707	ID2	NM_002166
563	chr1	15556115	15556239	EFHD2	NM_024329
564	chr22	29903552	29903667	RNF185	NM_152267
565	chr5	172262320	172262426	ERGIC1	NM_020462
566	chr5	79982913	79983290	MSH3	NM_002439
567	chr5	134288947	134289139	PCBD2	NM_032151
568	chr7	80409489	80409687	SEMA3C	NM_006379
569	chr2	74945625	74945818	HK2	NM_000189
570	chr1	158361116	158361270	ATP1A2	NM_000702
571	chr19	4664689	4664813	DPP9	NM_139159
572	chr7	5290920	5291041	SLC29A4	NM_153247
573	chr6	44150985	44151105	MRPL14	NM_032111
574	chr19	60678177	60678297	ZNF628	NM_033113
575	chr19	8332959	8333077	ANGPTL4	NM_001039667
576	chrX	48230931	48231046	FTSJ1	NM_177434
577	chr20	55713529	55713643	TMEPAI	NM_020182
578	chr10	4736541	4736650	AKR1CL2	NM_001040177
579	chr5	149829560	149829667	RPS14	NM_001025070
580	chr16	85326137	85326240	FOXL1	NM_005250
581	chr15	91153411	91153514	CHD2	NM_001042572
582	chr6	131998504	131998603	ENPP3	NM_005021
583	chr21	44023452	44023551	CSTB	NM_000100
584	chr5	138885783	138886017	UBE2D2	NM_003339
585	chr1	46563865	46564041	NSUN4	NM_199044
586	chr12	36829680	36829808	ALG10B	NM_001013620
587	chr2	216271644	216271765	FN1	NM_054034
588	chr5	119118670	119118787	LOC340069	NM_182761
589	chr5	149833016	149833130	RPS14	NM_001025070
590	chr5	156393876	156393987	HAVCR1	NM_012206
591	chr6	44152370	44152480	MRPL14	NM_032111
592	chr22	35191236	35191341	TXN2	NM_012473
593	chr1	230648320	230648422	SIPA1L2	NM_020808
594	chr19	12768015	12768114	JUNB	NM_002229
595	chr16	80236479	80236578	PLCG2	NM_002661
596	chr3	14469449	14469571	SLC6A6	NM_003043
597	chr17	68224068	68224217	SLC39A11	NM_139177
598	chr7	45258033	45258253	RAMP3	NM_005856
599	chr5	37846206	37846380	GDNF	NM_199234
600	chr3	23176448	23176565	UBE2E2	NM_152653
601	chr7	139017590	139017714	HIPK2	NM_022740
602	chr16	29938596	29938704	DOC2A	NM_003586
603	chr4	120441216	120441392	LOC401152	NM_001001701
604	chr5	134290521	134290725	PCBD2	NM_032151

605	chr9	127063529	127063634	GAPVD1	NM_015635
606	chr8	119027959	119028066	EXT1	NM_000127
607	chr1	119258649	119258756	TBX15	NM_152380
608	chr2	191419779	191419878	GLS	NM_014905

8.3 DNA Sequence of "Nici"

Putative coding sequence for the "Nici" transcript (Clone IMAGE: 610879). Sequence alignment revealed one potential splicing site (highlighted in grey). Reverse strand of chr 12:8014523-8007477, HG 18.

GCCAATTTGGGTAAATGCGTCCAAGAGGAAGAAATAAAAATGAAAAGACTGACAGTTTAAAAGCATGTGG
AGGGCTTAGAAGATAAAGACGAGTGCAAGAAAGGGAGAGCGCGGCCGGCCGCGGTGTCTCACGCCTGTC
ATCCTAGCACTTTGGGAGGCCGAAGAGATCGGATCACTTGAGGTCCGGAGTTCGAATCCAGCCTGGCCAA
CATGGTGAAACCCCGTCTCTACTAAAAATACAAAAAAATTAGCCGGACCTGGTGGCGGGCGTCTGTAATGC
CAGCTACTCGGGAGGCTGAGGCGGGAGAATCGCTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCCGAG
ATCGCGCCACTGCATTCCAGCCTGGGCAACAGACTGAGACCCTGCCTCAAAAGAAAAAAAGAAAAAAAAA
AAAGGAGAGCGCGTACTGTTGCGGAAGGGGCACTGCTCGAGAAACAGAGACGAAGGGATGATTTTCAAG
TAATCAAGACACCGAGAAAGACAAGATTGGGGATGGAAGAGCAGTAAGAATTAATGGAGACGTCCAAGA
ATTGTAGTTCGGGCGCGGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCTGAGATGGAGTTTCGCTC
TTGTTGCCAGGCTGGAGTACAGTGGCCTGATCTCGGCTCACTGCAACCTCTGCCTCCTGGGTTTAAGCGATT
CTCCTGCCTTAGCTTCCTCCTGAGTAGCTTGGATTACAGGCATGCGTCACCATGCCCGGCTAATATTGTATTT
TAATGGAGACGGGATTTCACCATGTTGGCCAGGCTGGTCTCGAAGTCCTGACCTCAAGTGATCCGCCCGCC
TCATCCTCTCAAAGTGCTGGGGTTACAGGCATTAGCCCTCACGCCTGGCCAAATGCAGTCTTTTCACCACCT
TTTTGGATCACAGAAAATGACCATGAAGCTATGGAAAAAAGGGAGAGTTGAATTTAAGATTAGCTGTCAG
CCCATCTTTCATTAGCCATTTCTCTAATCCAGTCTTGAATTTTCCTCTTTCTAGTAATATGGAGGGTAGCACTA
TGTTCAAAGCGGCCACTGGGGATTTAGGTATTTTAATATCTGTACTATGTTTCTAGATAAGTAAAATAACCA
GGTGAATTTATCAGCCCACAGTGAACCTGGGGAACAGTTTATTGAACTAGTGCCCTTAAGTAGAGGACATT
CTCAACATTGCTAAGTTTATCTCCCATTCCCTATATCATGGAATGAGAGAAATTTGGAGAAAAAATTTCTATT
TCCTAATCTCAGCTTTTTTTTTTTTTTTTTCCGAGACAGAGTATCGAGCTGTCGCCCAGGCTGGAGTGCAAT
GGCATGATCTTGGCTCACTGCAACCTCTGCCTCCCGCGTTCTAAGTGATTCTCCTTCCTCAGCCTCCCGAATG
GCTGGGATTATAGGCACGCACCATCATGCATGGCTAATTTTGTATTTTTAGTAGAGATGGGGTTTCCCCATG
TTGGCCAGGCTGGTCTCGAACTCCTGATCTCAGGTTATCCACCCGCCTCAGCCTCTCCAAGTGCTGGGATTA
CAAGTGTGAGCCACCGCACCCCGCCCTAATTTCAGCCTTTTTATATTTTCTTGGTAGGCCCACTTTTTGGCAT
CCCCTAACAAGAAGTCCTCAGAGATTTTTACTACTCAACATGTGTGCGTCCCACTCAGATGTATTGTGTTCTT
TCTGTTCACCCTTTACTTCCCATTCCATTTCTTCCTTTGACTTAGGGTTAGAATGGCCCAATGATCCTACAACT
TTTTGATGCTATTTCATTTGATTCCTCTACTGTCTAAATCACAAGTTGGCTAACTTTACTTACTGTAAGGGATC
AGATAGCAAATAGTTTAGGCCTTGTGGGCCATTCAATCTGTCATAACTACCCACTCTGACCATATGTAGTGC
AAAAGCAGCCCTTGACAATATATAAATGAATGAACGTGAATGTATTTCAATAAAGCTTTACTTGTGGACACT
GAAA

8.4 Side by side alignment of the "Nici" sequence derived from clone (Clone IMAGE: 610879) with the reference genome HG18

Block1

```
0000001 gccaatttgggtaaatgcgtccaagaggaagaaataaaaatgaaaagact 0000050
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8014523 gccaatttgggtaaatgcgtccaagaggaagaaataaaaatgaaaagact 8014474

0000051 gacagtttaaaagcatgtggagggcttagaagataaagacgagtgcaaga 0000100
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8014473 gacagtttaaaagcatgtggagggcttagaagataaagacgagtgcaaga 8014424

0000101 aagggagagcgcggccggccgcggtgtctcacgcctgtcatcctagcact 0000150
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8014423 aagggagagcgcggccggccgcggtgtctcacgcctgtcatcctagcact 8014374

0000151 ttgggaggccgaagagatcggatcacttgaggtccggagttcgaatccag 0000200
<<<<<<< |||||||||||||||||||||||||||||||||||||||||| ||||||| <<<<<<<
8014373 ttgggaggccgaagagatcggatcacttgaggtccggagttcaaatccag 8014324

0000201 cctggccaacatggtgaaaccccgtctctactaaaaatacaaaaaaatta 0000250
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8014323 cctggccaacatggtgaaaccccgtctctactaaaaatacaaaaaaatta 8014274

0000251 gccggacctggtggcgggcgtctgtaatgccagctactcgggaggctgag 0000300
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8014273 gccggacctggtggcgggcgtctgtaatgccagctactcgggaggctgag 8014224

0000301 gcgggagaatcgcttgaacccgggaggcggaggttgcagtgagccgagat 0000350
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8014223 gcgggagaatcgcttgaacccgggaggcggaggttgcagtgagccgagat 8014174

0000351 cgcgccactgcattccagcctgggcaacagactgagaccctgcctcaaaa 0000400
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8014173 cgcgccactgcattccagcctgggcaacagactgagaccctgcctcaaaa 8014124

0000401 gaaaaaaagaaaaaaaaaaaaggagagcgcgtactgttgcggaaggggca 0000450
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8014123 gaaaaaaagaaaaaaaaaaaaggagagcgcgtactgttgcggaaggggca 8014074

0000451 ctgctcgagaaacagagacgaagggatgattttcaagtaatcaagacacc 0000500
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8014073 ctgctcgagaaacagagacgaagggatgattttcaagtaatcaagacacc 8014024

0000501 gagaaagacaagattggggatggaagagcagtaagaattaatggagacgt 0000550
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8014023 gagaaagacaagattggggatggaagagcagtaagaattaatggagacgt 8013974

0000551 ccaagaattgtagttcgggcgcggtggctcatgcctgtaatcccagcact 0000600
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8013973 ccaagaattgtagttcgggcgcggtggctcatgcctgtaatcccagcact 8013924

0000601 ttgggaggctgag 0000613
<<<<<<< ||||||||||||| <<<<<<<
8013923 ttgggaggctgag 8013911
```

Block2

```
0000614 atggagtttcgctcttgttgccaggctggagtacagtggcctgatctcgg 0000663
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8008871 atggagtttcgctcttgttgccaggctggagtacagtggcctgatctcgg 8008822

0000664 ctcactgcaacctctgcctcctgggtttaagcgattctcctgccttagct 0000713
<<<<<<< |||||||||||||||||||||||||||||||| |||||||||||||||| <<<<<<<
8008821 ctcactgcaacctctgcctcctgggtttaagcaattctcctgccttagct 8008772

0000714 tcctcctgagtagcttggattacaggcatgcgtcaccatgcccggctaat 0000763
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8008771 tcctcctgagtagcttggattacaggcatgcgtcaccatgcccggctaat 8008722

0000764 attgtattttaatggagacgggatttcaccatgttggccaggctggtctc 0000813
<<<<<<< |||||||||||||| |||||||||||||||||||||||||||||||||| <<<<<<<
8008721 attgtattttaatagagacgggatttcaccatgttggccaggctggtctc 8008672

0000814 gaagtcctgacctcaagtgatccgcccgcctcatcctctcaaagtgctgg 0000863
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8008671 gaagtcctgacctcaagtgatccgcccgcctcatcctctcaaagtgctgg 8008622

0000864 ggttacaggcattagccctcacgcctggccaaatgcagtcttttcaccac 0000913
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8008621 ggttacaggcattagccctcacgcctggccaaatgcagtcttttcaccac 8008572

0000914 ctttttggatcacagaaaatgaccatgaagctatggaaaaaagggagagt 0000963
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8008571 ctttttggatcacagaaaatgaccatgaagctatggaaaaaagggagagt 8008522

0000964 tgaatttaagattagctgtcagcccatctttcattagccatttctctaat 0001013
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8008521 tgaatttaagattagctgtcagcccatctttcattagccatttctctaat 8008472

0001014 ccagtcttgaattttcctctttctagtaatatggagggtagcactatgtt 0001063
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8008471 ccagtcttgaattttcctctttctagtaatatggagggtagcactatgtt 8008422

0001064 caaagcggccactggggatttaggtattttaatatctgtactatgtttct 0001113
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8008421 caaagcggccactggggatttaggtattttaatatctgtactatgtttct 8008372

0001114 agataagtaaaataaccaggtgaatttatcagcccacagtgaacctgggg 0001163
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8008371 agataagtaaaataaccaggtgaatttatcagcccacagtgaacctgggg 8008322

0001164 aacagtttattgaactagtgcccttaagtagaggacattctcaacattgc 0001213
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8008321 aacagtttattgaactagtgcccttaagtagaggacattctcaacattgc 8008272

0001214 taagtttatctcccattccctatatcatggaatgagagaaatttggagaa 0001263
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8008271 taagtttatctcccattccctatatcatggaatgagagaaatttggagaa 8008222

0001264 aaaatttctatttcctaatctcagc.ttttttttttttttttttccgaga 0001312
<<<<<<< |||||||||||||||||||||||| |||||||||||||||||||||||| <<<<<<<
8008221 aaaatttctatttcctaatctcagctttttttttttttttttttccgaga 8008172
```

```
0001313 cagagtatcgagctgtcgcccaggctggagtgcaatggcatgatcttggc 0001362
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8008171 cagagtatcgagctgtcgcccaggctggagtgcaatggcatgatcttggc 8008122

0001363 tcactgcaacctctgcctcccgcgttctaagtgattctccttcctcagcc 0001412
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8008121 tcactgcaacctctgcctcccgcgttctaagtgattctccttcctcagcc 8008072

0001413 tcccgaatggctgggattataggcacgcaccatcatgcatggctaatttt 0001462
<<<<<<< |||||||| |||||||||||||||||||||||||||||||||||||||| <<<<<<<
8008071 tcccgaatagctgggattataggcacgcaccatcatgcatggctaatttt 8008022

0001463 gtatttttagtagagatggggtttccccatgttggccaggctggtctcga 0001512
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8008021 gtatttttagtagagatggggtttccccatgttggccaggctggtctcga 8007972

0001513 actcctgatctcaggttatccacccgcctcagcctctccaagtgctggga 0001562
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8007971 actcctgatctcaggttatccacccgcctcagcctctccaagtgctggga 8007922

0001563 ttacaagtgtgagccaccgcaccccgccctaatttcagccttttatatt 0001612
<<<<<<< |||||||||||||||||||||||||||||||| ||||||||||||||  <<<<<<<
8007921 ttacaagtgtgagccaccgcaccccgccctaatctcagccttttatatc 8007872

0001613 ttcttggtaggcccacttttggcatcccctaacaagaagtcctcagaga 0001662
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8007871 ttcttggtaggcccacttttggcatcccctaacaagaagtcctcagaga 8007822

0001663 tttttactactcaacatgtgtgcgtcccactcagatgtattgtgttcttt 0001712
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8007821 tttttactactcaacatgtgtgcgtcccactcagatgtattgtgttcttt 8007772

0001713 ctgttcacccttacttcccattccatttcttcctttgacttagggttag 0001762
<<<<<<< |||||||||||| |||||||||||||||||||||| |||||||||||||| <<<<<<<
8007771 ctgttcaccctctacttcccattccatttcttcctctgacttagggttag 8007722

0001763 aatggcccaatgatcctacaacttttgatgctatttcatttgattcctc 0001812
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8007721 aatggcccaatgatcctacaacttttgatgctatttcatttgattcctc 8007672

0001813 tactgtctaaatcacaagttggctaactttacttactgtaagggatcaga 0001862
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8007671 tactgtctaaatcacaagttggctaactttacttactgtaagggatcaga 8007622

0001863 tagcaaatagtttaggccttgtgggccattcaatctgtcataactaccca 0001912
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8007621 tagcaaatagtttaggccttgtgggccattcaatctgtcataactaccca 8007572

0001913 ctctgaccatatgtagtgcaaaagcagcccttgacaatatataaatgaat 0001962
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8007571 ctctgaccatatgtagtgcaaaagcagcccttgacaatatataaatgaat 8007522

0001963 gaacgtgaatgtatttcaataaagctttacttgtggacactgaaa 0002007
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
8007521 gaacgtgaatgtatttcaataaagctttacttgtggacactgaaa 8007477
```

8.5 ChIP-seq sequences of HIF-binding sites at 11q13.3

H

HG 18 chr11:68943716-68944005

CTGGGAAGACAGAGCTGGTTCTGCACGTACCTCCGCTCCACAGTCACGGACACTGAGGTGCTCAGGGGTC
GGCGCTCAGGATCACTGGGCAGCCGTACGTGTCCCAGGCAGCGGGAGCCCCGGCAGTGCCAGCTCCCACA
CCATCCGGCTGTGACAACCTCAGGTCCTTGATGACAGGGAGCCTCTGCCAAGACGGTCCGAAAATTGGGA
AGGGGACCCGCTCATCTTTGTGGGGTCCCCAAGAGCTTCAGATAAGTTTGCTGAGCACCGGCTTCCTGATC
ACATTTCAC

h1

HG 18 chr11:68944502-68944811

CCGAGGGGCCACCCATTGTCTCCCAACCTCGGTTCCCAGGACTCCCGTCCCCACGGGCCCTCTAGACTCTGT
GAGGCTGTGATTCTGGGCAGGGTGAAGCCGGACATGCCGCAGGGCCGATGCCGGCCGTTGTGGAACCAG
CCCTGGAATCTCAAAAATGTCCTGTGGTCAGGCCTGAGCTTTGTTAGTGACTCAGCAGAAAGCACCCAGCC
CTGGCCCCGTGGCAGGTAGGCCAGCCCCAACCTGGGGTGCAGCCTCAGGCCTCCAGCGTCCAGAGCCCAG
GCCTGTGGGGATGTGGGCCCAGTGGGTG

h2

HG 18 chr11:68960982-68961226

GACAGAGCCCTTCTTCCCCTCAATGAGGCCACTGACCTCTCCAGGTGAACGTTGTTGGTGCAGCAGGGCAT
GAACCAGCTCACGTGCCTGGCCCACTGGCCAGACCACAGCATGGCTCTAGCTGCCCTGTCCCTGGCTGCAT
TGCTCACAGCCGACAGGGGCAGACAGGTGTAGACACAGCCGCCCAGCCCTGCAGCCCGGCATCCCAGGAA
TGTGGTCAGCCACCCACACCACTGCCCCCACCC

8.6 Abbreviations

Oct-4	POU domain transcription factor OCT4
2-OG	2-oxoglutarate
3C	chromatin conformation capture assay
ADM	adrenomedullin
ADP	adenosine diphosphate
AhR	aryl hydrocarbon receptor
ALDOA	aldolase A
ANOVA	analysis of variance
AP-1	activator protein 1
ARNT	aryl hydrocarbon receptor nuclear translocator
ARRDC3	arrestin domain containing 3
ATP	adenosine triphosphate
BAC	bacterial artificial chromosome
bHLH	basic helix-loop-helix
CA9	carboanhydrase 9
CAD	c-terminal activation domain
CBP	CREB-binding protein
CCND1	cyclin D1
ccRCC	clear-cell renal cell carcinoma
cDNA	complementary DNA
ChIP	chromatin immunoprecipitation
ChIP-seq	chromatin immunoprecipitation-sequencing
c-MYC	avian myelocytomatosis viral oncogene homolog
CODDD	c-terminal oxygen dependent degradation domain
ct	cycle threshold
C-TAD	c-terminal activation domain
dHIF	drosophila HIF
DMOG	Dimethyloxalylglycine
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
DNAse1	deoxyribonuclease 1
dNTP	deoxyribonucleotide triphosphate
DTT	dithiothreitol
E-box	enhancer box
EDTA	ethylendiamine tetraacetic acid disodium salt
EEF1G	eukaryotic translation elongation factor 1 gamma
EGLN	egl nine-like protein
ENCODE	The Encyclopedia of DNA Elements
EPAS1	endothelial PAS domain protein 1
EPO	erythropoietin
ER	estrogen receptor
ERRFI1	ERBB receptor feedback inhibitor 1
ES	enrichment score

FAIRE	formaldehyde assisted isolation of regulatory elements
FAM13A1	family with sequence similarity 13, member A
Fe^{2+}	Iron (II)
FH	fumarate hydratase
FIH-1	factor inhibiting HIF-1
FISH	fluorescent in-situ hybridisation
FOSL2	FOS-like antigen 2
FOXP3	forkhead box P3
FPKM	fragments per kilobase of exon per million fragments mapped
GET	glucose EDTA TRIS solution
GLUT1	glucose transporter 1
GR	growth receptor
GSEA	gene set enrichment analysis
GWAS	genome-wide association study
GYS1	glycogen synthase 1
H3	histone 3
H3K27me3	histone 3 lysine 27 trimethylation
H3K36me3	histone 3 lysine 36 trimethylation
H3K4me1	histone 3 lysine 4 monomethylation
H3K4me2	histone 3 lysine 4 dimethylation
H3K4me3	histone 3 lysine 4 trimethylation
H3K79me3	histone 3 lysine 79 trimethylation
H3K9me2	histone 3 lysine 9 dimethylation
H3K9me3	histone 3 lysine 9 trimethylation
H4k20me3	histone 4 lysine 20 trimethylation
HCl	hydrogen chloride
HDM	histone demethylase
HERC2	HECT and RLD domain containing E3 ubiquitin protein ligase 2
HG18	human genome 18
HIF	hypoxia inducible factor
HIF1A	hypoxia inducible factor 1a
HIF1AN	hypoxia inducible factor 1, alpha subunit inhibitor
HIG2	hypoxia induced gene 2
HK1	hexokinase 1
HK2	hexokinase 2
Hotair	HOX antisense intergenic RNA
HOX	homeobox gene
HPRT	hypoxanthine guanine phosphoribosyl transferase
HRE	hypoxia responsive element
HRP	horse-radish peroxidase
HSS	hypersensitivity site
IB	immunoblot
JARID1A	jumonji, AT rich interactive domain 1A
JARID1B	jumonji, AT rich interactive domain 1B
JARID1C	jumonji, AT rich interactive domain 1C
JMJD1B	jmjC domain-containing histone demethylation protein 2B

kb	kilo base
KCl	potassium chloride
kPa	kilopascal
LB	Luria Broth
LD	linkage disequilibrium
LDHA	lactate dehydrogenase A
LiCl	lithium chloride
MA	moving average
MAX	MYC-associated factor X
Mb	mega base
MET	met proto-oncogene
MgCl	magnesium chloride
MLL5	myeloid/lymphoid or mixed-lineage leukemia 5 (trithorax homolog, Drosophila)
mM	milli molar
mRNA	messenger RNA
mRNA-seq	messenger RNA-sequencing
mTOR	mammalian target of rapamycin
MYEOV	myeloma overexpressed (in a subset of t(11;14) positive multiple myelomas)
NaCl	sodium chloride
NAD	n-terminal activation domain
NaHCO3	sodium bicarbonate
NaOH	sodium hydroxide
ncRNA	non-coding RNA
NDRG1	n-myc downregulated gene 1
NODDD	n-terminal oxygen dependent degradation domain
N-TAD	n-terminal transactivation domain
OCA2	oculocutaneous albinism II
ODDD	oxygen dependent degradation domain
p15	cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4)
p21	cyclin-dependent kinase inhibitor 1A (p21, Cip1)
p53	tumor protein p53
PAS	PER-ARNT-SIM
PBRM1	polybromo 1
PBS	phosphate buffered saline
PBS-T	phosphate buffered saline tween
PCR	polymerase chain reaction
PDGF	platelet-derived growth factor
PGK-1	phosphoglycerate kinase 1
PHD	prolyl hydroxylase domain-containing protein
PMSF	phenylmethanesulfonylfluoride
pO_2	oxygen partial pressure
PPIA	peptidylprolyl isomerase A (cyclophilin A)
pVHL	von Hippel-Lindau protein
qPCR	quantitative PCR

RCC	renal cell carcinoma
RNA	ribonucleic acid
RNA pol2	RNA polymerase 2
rpm	revolutions per minute
RT	reverse transcription
SAP30	Sin3A-associated protein, 30kDa
SDMG	site directed mutagenesis
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SET1	histone lysine methyltransferase Set1
SETD2	SET domain containing 2
SIM	single-minded
siRNA	silencing RNA
SLC2A3	solute carrier family 2 (facilitated glucose transporter), member 3
SNP	single nucleotide polymorphism
SOC	super optimal broth with catabolic repression medium
SP1	specificity protein 1
SWI/SNF	switch/sucrose nonfermentable chromatin remodelling proteins
TBE	TRIS/Borate/EDTA
TCF4	transcription factor 4
TE	TRIS-EDTA
TES	transcriptional end site
TFBS	transcription factor binding site
TGFA	transforming growth factor, alpha
TRIS	tris(hydroxymethyl)aminomethane
TSS	transcriptional start site
UTR	untranslated region
UTX	ubiquitously transcribed X chromosome tetratricopeptide repeat protein
VEGF	vascular endothelial growth factor
VHL	von Hippel-Lindau tumor suppressor, E3 ubiquitin protein ligase
XIST	X (inactive)-specific transcript (non-protein coding)

8.7 Publications arising from this thesis

Schödel J, Bardella C, Sciesielski LK, Brown JM, Pugh CW, Buckle V, Tomlinson IP, Ratcliffe PJ, Mole DR: Common genetic variants at the 11q13.3 renal cancer susceptibility locus influence binding of HIF to an enhancer of cyclin D1 expression. Nat Gen 2012, 44(4), 420-425

Schödel J, Oikonomopoulos S, Ragoussis J, Pugh CW, Ratcliffe PJ, Mole DR: High-resolution genome-wide mapping of HIF-binding sites by ChIP-seq. Blood 2011, 117(23):e207-217.

9 Bibliography

Adam, J., E. Hatipoglu, et al. (2011). "Renal cyst formation in Fh1-deficient mice is independent of the Hif/Phd pathway: roles for fumarate in KEAP1 succination and Nrf2 signaling." Cancer Cell **20**(4): 524-537.

Baba, M., S. Hirai, et al. (2003). "Loss of von Hippel-Lindau protein causes cell density dependent deregulation of CyclinD1 expression through hypoxia-inducible factor." Oncogene **22**(18): 2728-2738.

Bannister, A. J. and T. Kouzarides (2005). "Reversing histone methylation." Nature **436**(7054): 1103-1106.

Barski, A., S. Cuddapah, et al. (2007). "High-resolution profiling of histone methylations in the human genome." Cell **129**(4): 823-837.

Bartke, T., M. Vermeulen, et al. (2010). "Nucleosome-interacting proteins regulated by DNA and histone methylation." Cell **143**(3): 470-484.

Berner, R. A. (2009). "Phanerozoic atmospheric oxygen: New results using the GEOCARBSULF model." Am J Sci **309**(7): 603-606.

Berner, R. A., J. M. Vandenbrooks, et al. (2007). "Evolution. Oxygen and evolution." Science **316**(5824): 557-558.

Bertozzi, D., R. Iurlaro, et al. (2011). "Characterization of novel antisense HIF-1alpha transcripts in human cancers." Cell Cycle **10**(18): 3189-3197.

Beyer, S., M. M. Kristensen, et al. (2008). "The histone demethylases JMJD1A and JMJD2B are transcriptional targets of hypoxia-inducible factor HIF." J Biol Chem **283**(52): 36542-36552.

Bieda, M., X. Xu, et al. (2006). "Unbiased location analysis of E2F1-binding sites suggests a widespread role for E2F1 in the human genome." Genome Res **16**(5): 595-605.

Bindra, R. S., J. R. Vasselli, et al. (2002). "VHL-mediated hypoxia regulation of cyclin D1 in renal carcinoma cells." Cancer Res **62**(11): 3014-3019.

Bracken, C. P., A. O. Fedele, et al. (2006). "Cell-specific regulation of hypoxia-inducible factor (HIF)-1alpha and HIF-2alpha stabilization and transactivation in a graded oxygen environment." J Biol Chem **281**(32): 22575-22585.

Brookes, E. and A. Pombo (2009). "Modifications of RNA polymerase II are pivotal in regulating gene expression states." EMBO Rep **10**(11): 1213-1219.

Brown, C. J., A. Ballabio, et al. (1991). "A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome." Nature **349**(6304): 38-44.

Brown, C. J., R. G. Lafreniere, et al. (1991). "Localization of the X inactivation centre on the human X chromosome in Xq13." Nature **349**(6304): 82-84.

Brown, J., K. Saracoglu, et al. (2001). "Subtelomeric chromosome rearrangements are detected using an innovative 12-color FISH assay (M-TEL)." Nat Med **7**(4): 497-501.

Bruick, R. K. and S. L. McKnight (2001). "A conserved family of prolyl-4-hydroxylases that modify HIF." Science **294**(5545): 1337-1340.

Cao, Q., C. Qin, et al. (2012). "Chromosome 11q13.3 variant modifies renal cell cancer risk in a Chinese population." Mutagenesis **27**(3): 345-350.

Carroll, J. S., X. S. Liu, et al. (2005). "Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1." Cell **122**(1): 33-43.

Carroll, J. S., C. A. Meyer, et al. (2006). "Genome-wide analysis of estrogen receptor binding sites." Nat Genet **38**(11): 1289-1297.

Cavalier-Smith, T. (1992). "The number of symbiotic origins of organelles." Biosystems **28**(1-3): 91-106; discussion 107-108.

Chan, W. K., G. Yao, et al. (1999). "Cross-talk between the aryl hydrocarbon receptor and hypoxia inducible factor signaling pathways. Demonstration of competition and compensation." J Biol Chem **274**(17): 12115-12123.

Cockman, M. E., N. Masson, et al. (2000). "Hypoxia inducible factor-alpha binding and ubiquitylation by the von Hippel-Lindau tumor suppressor protein." J Biol Chem **275**(33): 25733-25741.

Compernolle, 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.

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.

Creyghton, M. P., A. W. Cheng, et al. (2010). "Histone H3K27ac separates active from poised enhancers and predicts developmental state." Proc Natl Acad Sci U S A **107**(50): 21931-21936.

Dalgliesh, G. L., K. Furge, et al. (2010). "Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes." Nature **463**(7279): 360-363.

De Santa, F., I. Barozzi, et al. (2010). "A large fraction of extragenic RNA pol II transcription sites overlap enhancers." PLoS Biol **8**(5): e1000384.

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.

Dole, M. (1965). "The natural history of oxygen." J Gen Physiol **49**(1): Suppl:5-27.

Dunham, I., A. Kundaje, et al. (2012). "An integrated encyclopedia of DNA elements in the human genome." Nature **489**(7414): 57-74.

Dyson, A. and M. Singer (2011). "Tissue oxygen tension monitoring: will it fill the void?" Curr Opin Crit Care **17**(3): 281-289.

Elvidge, G. P., L. Glenny, et al. (2006). "Concordant regulation of gene expression by hypoxia and 2-oxoglutarate-dependent dioxygenase inhibition: the role of HIF-1alpha, HIF-2alpha, and other pathways." J Biol Chem **281**(22): 15215-15226.

Ema, M., S. Taya, et al. (1997). "A novel bHLH-PAS factor with close sequence similarity to hypoxia-inducible factor 1alpha regulates the VEGF expression and is potentially involved in lung and vascular development." Proc Natl Acad Sci U S A **94**(9): 4273-4278.

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.

Esteller, M. (2011). "Non-coding RNAs in human disease." Nat Rev Genet **12**(12): 861-874.

Ferlay, J., H. R. Shin, et al. (2010). "Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008." Int J Cancer **127**(12): 2893-2917.

Finley, D. S., A. J. Pantuck, et al. (2011). "Tumor biology and prognostic factors in renal cell carcinoma." Oncologist **16 Suppl 2**: 4-13.

Firth, J. D., B. L. Ebert, et al. (1994). "Oxygen-regulated control elements in the phosphoglycerate kinase 1 and lactate dehydrogenase A genes: similarities with the erythropoietin 3' enhancer." Proc Natl Acad Sci U S A **91**(14): 6496-6500.

Flamme, I., T. Frohlich, et al. (1997). "HRF, a putative basic helix-loop-helix-PAS-domain transcription factor is closely related to hypoxia-inducible factor-1 alpha and developmentally expressed in blood vessels." Mech Dev **63**(1): 51-60.

Forsythe, J. A., B. H. Jiang, et al. (1996). "Activation of vascular endothelial growth factor gene transcription by hypoxia-inducible factor 1." Mol Cell Biol **16**(9): 4604-4613.

Frew, I. J. and W. Krek (2007). "Multitasking by pVHL in tumour suppression." Curr Opin Cell Biol **19**(6): 685-690.

Fullwood, M. J., M. H. Liu, et al. (2009). "An oestrogen-receptor-alpha-bound human chromatin interactome." Nature **462**(7269): 58-64.

Giannakakis, A., R. Sandaltzopoulos, et al. (2008). "miR-210 links hypoxia with cell cycle regulation and is deleted in human epithelial ovarian cancer." Cancer Biol Ther **7**(2): 255-264.

Gimm, T., M. Wiese, et al. (2010). "Hypoxia-inducible protein 2 is a novel lipid droplet protein and a specific target gene of hypoxia-inducible factor-1." FASEB J **24**(11): 4443-4458.

Giresi, P. G., J. Kim, et al. (2007). "FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin." Genome Res **17**(6): 877-885.

Giresi, P. G. and J. D. Lieb (2009). "Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements)." Methods **48**(3): 233-239.

Gnarra, J. R., K. Tory, et al. (1994). "Mutations of the VHL tumour suppressor gene in renal carcinoma." Nat Genet **7**(1): 85-90.

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., P. Lal, et al. (2008). "HIF-alpha effects on c-Myc distinguish two subtypes of sporadic VHL-deficient clear cell renal carcinoma." Cancer Cell **14**(6): 435-446.

Grober, O. M., M. Mutarelli, et al. (2011). "Global analysis of estrogen receptor beta binding to breast cancer cell genome reveals an extensive interplay with estrogen receptor alpha for target gene regulation." BMC Genomics **12**: 36.

Gu, Y. Z., S. M. Moran, et al. (1998). "Molecular characterization and chromosomal localization of a third alpha-class hypoxia inducible factor subunit, HIF3alpha." Gene Expr **7**(3): 205-213.

Guenther, M. G., S. S. Levine, et al. (2007). "A chromatin landmark and transcription initiation at most promoters in human cells." Cell **130**(1): 77-88.

Gumz, M. L., H. Zou, et al. (2007). "Secreted frizzled-related protein 1 loss contributes to tumor phenotype of clear cell renal cell carcinoma." Clin Cancer Res **13**(16): 4740-4749.

Guo, G., Y. Gui, et al. (2012). "Frequent mutations of genes encoding ubiquitin-mediated proteolysis pathway components in clear cell renal cell carcinoma." Nat Genet **44**(1): 17-19.

Gupta, R. A., N. Shah, et al. (2010). "Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis." Nature **464**(7291): 1071-1076.

Guttman, M., I. Amit, et al. (2009). "Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals." Nature **458**(7235): 223-227.

Hagege, H., P. Klous, et al. (2007). "Quantitative analysis of chromosome conformation capture assays (3C-qPCR)." Nat Protoc **2**(7): 1722-1733.

Haiman, C. A., N. Patterson, et al. (2007). "Multiple regions within 8q24 independently affect risk for prostate cancer." Nat Genet **39**(5): 638-644.

Hampton-Smith, R. J. and D. J. Peet (2009). "From polyps to people: a highly familiar response to hypoxia." Ann N Y Acad Sci **1177**: 19-29.

Harrison, J. F. and G. G. Haddad (2011). "Effects of oxygen on growth and size: synthesis of molecular, organismal, and evolutionary studies with Drosophila melanogaster." Annu Rev Physiol **73**: 95-113.

Heintzman, N. D., G. C. Hon, et al. (2009). "Histone modifications at human enhancers reflect global cell-type-specific gene expression." Nature **459**(7243): 108-112.

Heintzman, N. D., R. K. Stuart, et al. (2007). "Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome." Nat Genet **39**(3): 311-318.

Herman, J. G., F. Latif, et al. (1994). "Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma." Proc Natl Acad Sci U S A **91**(21): 9700-9704.

Higgs, D. R., J. D. Engel, et al. (2012). "Thalassaemia." Lancet **379**(9813): 373-383.

Hon, W. C., M. I. Wilson, et al. (2002). "Structural basis for the recognition of hydroxyproline in HIF-1 alpha by pVHL." Nature **417**(6892): 975-978.

Hu, C. J., S. Iyer, et al. (2006). "Differential regulation of the transcriptional activities of hypoxia-inducible factor 1 alpha (HIF-1alpha) and HIF-2alpha in stem cells." Mol Cell Biol **26**(9): 3514-3526.

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.

Huarte, M., M. Guttman, et al. (2010). "A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response." Cell **142**(3): 409-419.

Hutson, T. E. (2011). "Targeted therapies for the treatment of metastatic renal cell carcinoma: clinical evidence." Oncologist **16 Suppl 2**: 14-22.

Isaacs, J. S., Y. J. Jung, et al. (2005). "HIF overexpression correlates with biallelic loss of fumarate hydratase in renal cancer: novel role of fumarate in regulation of HIF stability." Cancer Cell **8**(2): 143-153.

Ivan, M., K. Kondo, et al. (2001). "HIFalpha targeted for VHL-mediated destruction by proline hydroxylation: implications for O2 sensing." Science **292**(5516): 464-468.

Iyer, N. V., L. E. Kotch, et al. (1998). "Cellular and developmental control of O2 homeostasis by hypoxia-inducible factor 1 alpha." Genes Dev **12**(2): 149-162.

Jaakkola, P., D. R. Mole, et al. (2001). "Targeting of HIF-alpha to the von Hippel-Lindau ubiquitylation complex by O2-regulated prolyl hydroxylation." Science **292**(5516): 468-472.

Jelkmann, W. and C. Bauer (1981). "Demonstration of high levels of erythropoietin in rat kidneys following hypoxic hypoxia." Pflugers Arch **392**(1): 34-39.

John, S., P. J. Sabo, et al. (2011). "Chromatin accessibility pre-determines glucocorticoid receptor binding patterns." Nat Genet **43**(3): 264-268.

Johnson, A. B., N. Denko, et al. (2008). "Hypoxia induces a novel signature of chromatin modifications and global repression of transcription." Mutat Res **640**(1-2): 174-179.

Kaelin, W. G. (2007). "Von Hippel-Lindau disease." Annu Rev Pathol **2**: 145-173.

Kaelin, W. G., Jr. (2009). "Treatment of kidney cancer: insights provided by the VHL tumor-suppressor protein." Cancer **115**(10 Suppl): 2262-2272.

Kaelin, W. G., Jr. and P. J. Ratcliffe (2008). "Oxygen sensing by metazoans: the central role of the HIF hydroxylase pathway." Mol Cell **30**(4): 393-402.

Khan, M. N., T. Bhattacharyya, et al. (2011). "Factor inhibiting HIF (FIH-1) promotes renal cancer cell survival by protecting cells from HIF-1alpha-mediated apoptosis." Br J Cancer **104**(7): 1151-1159.

Knauth, K., C. Bex, et al. (2006). "Renal cell carcinoma risk in type 2 von Hippel-Lindau disease correlates with defects in pVHL stability and HIF-1alpha interactions." Oncogene **25**(3): 370-377.

Kondo, K., W. Y. Kim, et al. (2003). "Inhibition of HIF2alpha is sufficient to suppress pVHL-defective tumor growth." PLoS Biol **1**(3): E83.

Kondo, K., J. Klco, et al. (2002). "Inhibition of HIF is necessary for tumor suppression by the von Hippel-Lindau protein." Cancer Cell **1**(3): 237-246.

Kouzarides, T. (2007). "Chromatin modifications and their function." Cell **128**(4): 693-705.

Krieg, A. J., E. B. Rankin, et al. (2010). "Regulation of the histone demethylase JMJD1A by hypoxia-inducible factor 1 alpha enhances hypoxic gene expression and tumor growth." Mol Cell Biol **30**(1): 344-353.

Lando, D., D. J. Peet, et al. (2002). "Asparagine hydroxylation of the HIF transactivation domain a hypoxic switch." Science **295**(5556): 858-861.

Larkin, J., X. Y. Goh, et al. (2012). "Epigenetic regulation in RCC: opportunities for therapeutic intervention?" Nat Rev Urol **9**(3): 147-155.

Lau, K. W., Y. M. Tian, et al. (2007). "Target gene selectivity of hypoxia-inducible factor-alpha in renal cancer cells is conveyed by post-DNA-binding mechanisms." Br J Cancer **96**(8): 1284-1292.

Li, L., L. Zhang, et al. (2007). "Hypoxia-inducible factor linked to differential kidney cancer risk seen with type 2A and type 2B VHL mutations." Mol Cell Biol **27**(15): 5381-5392.

Loenarz, C., M. L. Coleman, et al. (2011). "The hypoxia-inducible transcription factor pathway regulates oxygen sensing in the simplest animal, Trichoplax adhaerens." EMBO Rep **12**(1): 63-70.

Maltepe, E., J. V. Schmidt, et al. (1997). "Abnormal angiogenesis and responses to glucose and oxygen deprivation in mice lacking the protein ARNT." Nature **386**(6623): 403-407.

Mandriota, S. J., K. J. Turner, et al. (2002). "HIF activation identifies early lesions in VHL kidneys: evidence for site-specific tumor suppressor function in the nephron." Cancer Cell **1**(5): 459-468.

Margulis, L. and D. Bermudes (1985). "Symbiosis as a mechanism of evolution: status of cell symbiosis theory." Symbiosis **1**: 101-124.

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.

Maynard, M. A., A. J. Evans, et al. (2005). "Human HIF-3alpha4 is a dominant-negative regulator of HIF-1 and is down-regulated in renal cell carcinoma." FASEB J **19**(11): 1396-1406.

Maynard, M. A., H. Qi, et al. (2003). "Multiple splice variants of the human HIF-3 alpha locus are targets of the von Hippel-Lindau E3 ubiquitin ligase complex." J Biol Chem **278**(13): 11032-11040.

McIntosh, B. E., J. B. Hogenesch, et al. (2010). "Mammalian Per-Arnt-Sim proteins in environmental adaptation." Annu Rev Physiol **72**: 625-645.

Miescher, F. (1893). "Uber die Beziehungen zwischen Meereshohe und Beschaffenheit des Blutes." Correspondenz-Blatt Schweizer Arzte(23): 809.

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.

Mortazavi, A., B. A. Williams, et al. (2008). "Mapping and quantifying mammalian transcriptomes by RNA-Seq." Nat Methods **5**(7): 621-628.

Musgrove, E. A., C. E. Caldon, et al. (2011). "Cyclin D as a therapeutic target in cancer." Nat Rev Cancer **11**(8): 558-572.

Ng, H. H., F. Robert, et al. (2003). "Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity." Mol Cell **11**(3): 709-719.

Nowick, K. and L. Stubbs (2010). "Lineage-specific transcription factors and the evolution of gene regulatory networks." Brief Funct Genomics **9**(1): 65-78.

Nytko, K. J., P. Spielmann, et al. (2007). "Regulated function of the prolyl-4-hydroxylase domain (PHD) oxygen sensor proteins." Antioxid Redox Signal **9**(9): 1329-1338.

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.

Park, P. J. (2009). "ChIP-seq: advantages and challenges of a maturing technology." Nat Rev Genet **10**(10): 669-680.

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.

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.

Pollard, P. J., C. Loenarz, et al. (2008). "Regulation of Jumonji-domain-containing histone demethylases by hypoxia-inducible factor (HIF)-1alpha." Biochem J **416**(3): 387-394.

Pomerantz, M. M., N. Ahmadiyeh, et al. (2009). "The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer." Nat Genet **41**(8): 882-884.

Purdue, M. P., M. Johansson, et al. (2011). "Genome-wide association study of renal cell carcinoma identifies two susceptibility loci on 2p21 and 11q13.3." Nat Genet **43**(1): 60-65.

Rada-Iglesias, A., R. Bajpai, et al. (2011). "A unique chromatin signature uncovers early developmental enhancers in humans." Nature **470**(7333): 279-283.

Ramachandrareddy, H., A. Bouska, et al. (2010). "BCL6 promoter interacts with far upstream sequences with greatly enhanced activating histone modifications in germinal center B cells." Proc Natl Acad Sci U S A **107**(26): 11930-11935.

Raval, R. R., K. W. Lau, et al. (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**(13): 5675-5686.

Rinn, J. L. and H. Y. Chang (2012). "Genome regulation by long noncoding RNAs." Annu Rev Biochem **81**: 145-166.

Rinn, J. L., M. Kertesz, et al. (2007). "Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs." Cell **129**(7): 1311-1323.

Rosenberger, C., S. Mandriota, et al. (2002). "Expression of hypoxia-inducible factor-1alpha and -2alpha in hypoxic and ischemic rat kidneys." J Am Soc Nephrol **13**(7): 1721-1732.

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.

Rytkonen, K. T., T. A. Williams, et al. (2011). "Molecular evolution of the metazoan PHD-HIF oxygen-sensing system." Mol Biol Evol **28**(6): 1913-1926.

Schito, L., S. Rey, et al. (2012). "Hypoxia-inducible factor 1-dependent expression of platelet-derived growth factor B promotes lymphatic metastasis of hypoxic breast cancer cells." Proc Natl Acad Sci U S A **109**(40): E2707-2716.

Schodel, J., D. Bohr, et al. (2010). "Factor inhibiting HIF limits the expression of hypoxia-inducible genes in podocytes and distal tubular cells." Kidney Int **78**(9): 857-867.

Schodel, J., S. Oikonomopoulos, et al. (2011). "High-resolution genome-wide mapping of HIF-binding sites by ChIP-seq." Blood **117**(23): e207-217.

Schoenfelder, S., I. Clay, et al. (2010). "The transcriptional interactome: gene expression in 3D." Curr Opin Genet Dev **20**(2): 127-133.

Scortegagna, M., K. Ding, et al. (2003). "Multiple organ pathology, metabolic abnormalities and impaired homeostasis of reactive oxygen species in Epas1-/- mice." Nat Genet **35**(4): 331-340.

Semenza, G. L. (2007). "Life with oxygen." Science **318**(5847): 62-64.

Semenza, G. L. (2012). "Hypoxia-inducible factors in physiology and medicine." Cell **148**(3): 399-408.

Semenza, G. L., M. K. Nejfelt, et al. (1991). "Hypoxia-inducible nuclear factors bind to an enhancer element located 3' to the human erythropoietin gene." Proc Natl Acad Sci U S A **88**(13): 5680-5684.

Semenza, G. L., P. H. Roth, et al. (1994). "Transcriptional regulation of genes encoding glycolytic enzymes by hypoxia-inducible factor 1." J Biol Chem **269**(38): 23757-23763.

Shen, C., R. Beroukhim, et al. (2011). "Genetic and Functional Studies Implicate HIF1alpha as a 14q Kidney Cancer Suppressor Gene." Cancer Discov **1**(3): 222-235.

Sogawa, K., R. Nakano, et al. (1995). "Possible function of Ah receptor nuclear translocator (Arnt) homodimer in transcriptional regulation." Proc Natl Acad Sci U S A **92**(6): 1936-1940.

Swanson, H. I. (2002). "DNA binding and protein interactions of the AHR/ARNT heterodimer that facilitate gene activation." Chem Biol Interact **141**(1-2): 63-76.

Swanson, H. I. and J. H. Yang (1999). "Specificity of DNA binding of the c-Myc/Max and ARNT/ARNT dimers at the CACGTG recognition site." Nucleic Acids Res **27**(15): 3205-3212.

Swedenborg, E. and I. Pongratz (2010). "AhR and ARNT modulate ER signaling." Toxicology **268**(3): 132-138.

Swiezewski, S., F. Liu, et al. (2009). "Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target." Nature **462**(7274): 799-802.

Takeda, K., A. Cowan, et al. (2007). "Essential role for prolyl hydroxylase domain protein 2 in oxygen homeostasis of the adult vascular system." Circulation **116**(7): 774-781.

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.

Tan, C. C., K. U. Eckardt, et al. (1992). "Feedback modulation of renal and hepatic erythropoietin mRNA in response to graded anemia and hypoxia." Am J Physiol **263**(3 Pt 2): F474-481.

Tan, M., H. Luo, et al. (2011). "Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification." Cell **146**(6): 1016-1028.

Tanaka, T., M. Wiesener, et al. (2009). "The human HIF (hypoxia-inducible factor)-3alpha gene is a HIF-1 target gene and may modulate hypoxic gene induction." Biochem J **424**(1): 143-151.

Tanimoto, K., K. Tsuchihara, et al. (2010). "Genome-wide identification and annotation of HIF-1alpha binding sites in two cell lines using massively parallel sequencing." Hugo J **4**(1-4): 35-48.

Taylor, C. T. and J. C. McElwain (2010). "Ancient atmospheres and the evolution of oxygen sensing via the hypoxia-inducible factor in metazoans." Physiology (Bethesda) **25**(5): 272-279.

Tchou-Wong, K. M., K. Kiok, et al. (2011). "Effects of nickel treatment on H3K4 trimethylation and gene expression." PLoS One **6**(3): e17728.

Thrash-Bingham, C. A. and K. D. Tartof (1999). "aHIF: a natural antisense transcript overexpressed in human renal cancer and during hypoxia." J Natl Cancer Inst **91**(2): 143-151.

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.

Tomlinson, I., E. Webb, et al. (2007). "A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21." Nat Genet **39**(8): 984-988.

Tomlinson, I. P., N. A. Alam, et al. (2002). "Germline mutations in FH predispose to dominantly inherited uterine fibroids, skin leiomyomata and papillary renal cell cancer." Nat Genet **30**(4): 406-410.

Tuupanen, S., M. Turunen, et al. (2009). "The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling." Nat Genet **41**(8): 885-890.

Uchida, T., F. Rossignol, et al. (2004). "Prolonged hypoxia differentially regulates hypoxia-inducible factor (HIF)-1alpha and HIF-2alpha expression in lung epithelial cells: implication of natural antisense HIF-1alpha." J Biol Chem **279**(15): 14871-14878.

van Haaften, G., G. L. Dalgliesh, et al. (2009). "Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer." Nat Genet **41**(5): 521-523.

Varela, I., P. Tarpey, et al. (2011). "Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma." Nature **469**(7331): 539-542.

Viault, F. (1890). "Sur l'augmentation considérable du nombre des globules rouges dans le sang chez les habitants des hauts plateaux de 1'Amérique du Sud." Comptes Rendus Hebdomaires des Seances de l'Académie des Sciences **111**: 917–918.

Visser, M., M. Kayser, et al. (2012). "HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter." Genome Res **22**(3): 446-455.

Wang, G. L. and G. L. Semenza (1993). "General involvement of hypoxia-inducible factor 1 in transcriptional response to hypoxia." Proc Natl Acad Sci U S A **90**(9): 4304-4308.

Wang, G. L. and G. L. Semenza (1995). "Purification and characterization of hypoxia-inducible factor 1." J Biol Chem **270**(3): 1230-1237.

Warnecke, C., A. Weidemann, et al. (2008). "The specific contribution of hypoxia-inducible factor-2alpha to hypoxic gene expression in vitro is limited and modulated by cell type-specific and exogenous factors." Exp Cell Res **314**(10): 2016-2027.

Warnecke, C., Z. Zaborowska, et al. (2004). "Differentiating the functional role of hypoxia-inducible factor (HIF)-1alpha and HIF-2alpha (EPAS-1) by the use of RNA interference: erythropoietin is a HIF-2alpha target gene in Hep3B and Kelly cells." FASEB J **18**(12): 1462-1464.

Wei, C. L., Q. Wu, et al. (2006). "A global map of p53 transcription-factor binding sites in the human genome." Cell **124**(1): 207-219.

Welboren, W. J., M. A. van Driel, et al. (2009). "ChIP-Seq of ERalpha and RNA polymerase II defines genes differentially responding to ligands." EMBO J **28**(10): 1418-1428.

Wellmann, S., M. Bettkober, et al. (2008). "Hypoxia upregulates the histone demethylase JMJD1A via HIF-1." Biochem Biophys Res Commun **372**(4): 892-897.

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., J. S. Jurgensen, et al. (2003). "Widespread hypoxia-inducible expression of HIF-2alpha in distinct cell populations of different organs." FASEB J **17**(2): 271-273.

Wiesener, M. S., P. H. Maxwell, et al. (2009). "Novel insights into the role of the tumor suppressor von Hippel Lindau in cellular differentiation, ciliary biology, and cyst repression." J Mol Med (Berl) **87**(9): 871-877.

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.

Wu, X., G. Scelo, et al. (2012). "A genome-wide association study identifies a novel susceptibility locus for renal cell carcinoma on 12p11.23." Hum Mol Genet **21**(2): 456-462.

Wutz, A. (2011). "Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation." Nat Rev Genet **12**(8): 542-553.

Wykoff, C. C., C. Sotiriou, et al. (2004). "Gene array of VHL mutation and hypoxia shows novel hypoxia-induced genes and that cyclin D1 is a VHL target gene." Br J Cancer **90**(6): 1235-1243.

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

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.

Xu, J., B. Wang, et al. (2012). "Epigenetic regulation of HIF-1alpha in renal cancer cells involves HIF-1alpha/2alpha binding to a reverse hypoxia-response element." Oncogene **31**(8): 1065-1072.

Young, A. P., S. Schlisio, et al. (2008). "VHL loss actuates a HIF-independent senescence programme mediated by Rb and p400." Nat Cell Biol **10**(3): 361-369.

Yu, W., D. Gius, et al. (2008). "Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA." Nature **451**(7175): 202-206.

Zanke, B. W., C. M. Greenwood, et al. (2007). "Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24." Nat Genet **39**(8): 989-994.

Zatyka, M., N. F. da Silva, et al. (2002). "Identification of cyclin D1 and other novel targets for the von Hippel-Lindau tumor suppressor gene by expression array analysis and investigation of cyclin D1 genotype as a modifier in von Hippel-Lindau disease." Cancer Res **62**(13): 3803-3811.

Zheng, Y., S. Z. Josefowicz, et al. (2007). "Genome-wide analysis of Foxp3 target genes in developing and mature regulatory T cells." Nature **445**(7130): 936-940.

Zhou, X., H. Sun, et al. (2010). "Hypoxia induces trimethylated H3 lysine 4 by inhibition of JARID1A demethylase." Cancer Res **70**(10): 4214-4221.

10 Declaration

Except where acknowledgement is made, the work described in this thesis was performed by the author in the Centre for Cellular and Molecular Physiology, University of Oxford. The work reported in this thesis has not been submitted for any other degree in this or any other university or institute of learning.

11 Acknowledgments

I am grateful to Prof. Peter Ratcliffe who in his position as Head of Department and supervisor gave me the opportunity to conduct my Dphil studies in his laboratory. I would also like to thank him for the many helpful scientific and philosophical discussions about hypoxia and life itself. My special gratitude goes to Dr. David Mole who as the direct supervisor guided my work with great endurance and competence. I would like to thank Prof. Chris Pugh and Prof. Chris Schofield for their substantial input and helpful discussions, and Prof. Doug Higgs and Prof. Peter Robbins for a stimulating discussion during the transfer viva.

My thanks go to Dr. Mat Coleman and Dr. Matt Cockman for the eminent support in technical and social issues and the introduction to big-game fishing, golfing and the culture of the British coast.

I would like to thank Dr. Norma Masson and Dr. Atsushi Yamamoto for their help in solving technical problems and providing useful material and tools, Melroy Xavier Miranda, Adam Zayer and Dr. Paddy Pollard for introducing me to Indian, Swedish and English culture, Prof. Terry Lappin for his constant humour and Dr. Lina Sciesielski for her professional support in technical problems.

My gratitude also goes to Rachelle, Linda, Ya min, Julie, Penny, Phebee, James, Catherine, Lynn, Tammie and Samira for giving me a great time in and outside the lab.

I dedicate this work to Nicoletta and Lieselotte, to my brother Christian and his family and to my parents, who always supported me.