

**Characterising the determinants of
hypoxia inducible transcription factors
binding to chromatin**



James Smythies

St. Hilda's College

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2017

Abstract

Hypoxia regulates many hundreds of genes that play important roles in numerous physiological and pathophysiological processes. The hypoxia inducible transcription factors (HIFs) are central to the transcriptional activation of these hypoxia-regulated genes. However, to date, little is known about the determinants of HIF-1 and HIF-2 binding site selection. Both HIF-1 and HIF-2 appear to bind as HIF- α /HIF-1 β heterodimers, and recognise the same core consensus DNA motif, the hypoxia response element (HRE). However, each has its own distinct but partially overlapping set of binding sites that accounts for only a subset of the total accessible HRE motifs. Here, I have utilised ChIP-seq to systematically compare pan-genomic HIF-1 α , HIF-2 α and HIF-1 β DNA binding in multiple cell lines and have related this to RNA-seq analyses and other publically-accessible next-generation sequencing datasets. I show that endogenous HIF- α subunits exhibit a high-degree of binding site concordance with HIF-1 β , consistent with largely canonical binding of intact heterodimers. Despite cell-type specific differences in HIF-1 and HIF-2 binding site occupancy, each isoform exhibits a remarkable rigidity in its preference to bind either promoter-proximal or promoter-distal sites, respectively. These specific distribution patterns are unaffected by the absence of the other HIF- α subunit, suggesting that they do not result from exclusion of one isoform by competition with the other, but rather are discrete properties of each. Furthermore, hypoxia regulated genes neighbouring sites that are shared by both HIF-1 and HIF-2 are more likely to be regulated by HIF-1 when the site is closer to the gene and by HIF-2 when further away, indicating that post-binding mechanisms of transcriptional regulation also follow a similar pattern. Comparison of sites preferentially bound by HIF-1 and HIF-2, respectively, revealed associations with distinct histone environments, distinct accessory transcription factor binding motifs and distinct patterns of transcription factor binding site occupancy, suggesting that each may be influencing specific HIF-1 and HIF-2 binding site selection. In particular, both the AP-1 motif and AP-1 binding site occupancy were enriched within HIF-2 binding sites compared to HIF-1 sites. Intervention on AP-1 DNA-binding using the dominant-negative protein, AFos, attenuated HIF binding, specifically at sites co-occupied by AP-1 and HIF. This indicates that a cooperative relationship exists between the two transcription factors. However, binding of both HIF-1 α and HIF-2 α were affected suggesting that while AP-1 binding may account for the ability of HIF to bind some HRE motifs but not others, it is not a determinant of differential binding between the two isoforms. Overall, this work reveals remarkably distinct and functionally relevant patterns of HIF-1 and HIF-2 binding across the genome, and provides insight into underlying mechanisms of binding.

Table of contents

Chapter 1: Introduction	9
1.1 The hypoxia inducible transcription factor pathway	10
1.1.1 Oxygen, and the evolution of multicellular organisms	10
1.1.2 Hypoxia inducible transcription factors and the cellular oxygen-sensing pathway	11
1.2 HIF-1 and HIF-2 function <i>in vivo</i>	16
1.2.1 The role of HIF in development	16
1.2.2 Physiological and pathophysiological roles of HIF	17
1.2.3 The role of HIF in cancer	18
1.3 Pan-genomic approaches to studying HIF	21
1.3.1 Regulation of gene expression by HIF	21
1.3.2 Pan-genomic HIF DNA binding studies	23
1.3.3 HIF and chromatin conformation	26
1.4 Transcription factor binding site selection	27
1.4.1 Transcription factor cooperativity	27
1.4.2 Hypoxia, histone modifications and DNA methylation	31
1.5 Aims of this study	32
 Chapter 2: Understanding the concordance of HIF-α/β binding to chromatin, and identification of any instances of non-canonical HIF-α binding	 35
2.1 Introduction	36
2.2 Results	40
2.2.1 Assessment of concordance between ChIP-seq biological replicates	40
2.2.2 Peak calling concordance, and statistical assessment of ChIP-seq signal concordance between HIF- α and HIF-1 β	43
2.2.3 Re-assessment of published endogenous and exogenous HIF-1 α and HIF-2 α ChIP-seq data in 786O cells, and association of these data with matched analyses of HIF-1 α and HIF-2 α ChIP-seq in other cell lines	58
2.2.4 Further corroboration of the concordant relationship between endogenous HIF- α and HIF-1 β , with the exception of a small number of HIF-2 α binding sites	65

2.2.5 Exploring HIF-1 β binding sites for absence of HIF- α	72
2.3 Discussion	76
2.3.1 HIF- α and HIF-1 β are largely concordant genome-wide	76
2.3.2 Endogenous HIF ChIP-seq in HepG2, RCC4, HKC8 and 786O, and concordance of HIF- α and HIF-1 β	78
 Chapter 3: Identifying characteristic properties of <i>cis</i>-regulatory elements preferentially bound by either HIF-1α or HIF-2α.....	
3.1 Introduction	82
3.2 Results	85
3.2.1 HIF binding site conservation	85
3.2.2 Stratification and characterisation of sites preferentially bound by either HIF-1 or HIF-2.....	95
3.2.3 Association of sites preferentially bound by HIF-1 or HIF-2 with ENCODE transcription factor ChIP-seq data	113
3.2.4 Association of sites preferentially bound by HIF-1 or HIF-2 with ENCODE histone modification ChIP-seq data	122
3.3 Discussion	126
3.3.1 Cell-type specificity	126
3.3.2 Pan-genomic distributions of sites preferentially bound by HIF-1 or HIF- 2	127
3.3.3 Outside of the HRE, distinct sets of transcription factor binding motifs enrich within HIF-1 and HIF-2 sites, respectively	129
3.3.4 Distinct sets of transcription factor binding sites coincide with HIF-1 and HIF-2 binding sites	134
3.3.5 Contrasting histone environments at HIF-1 and HIF-2 sites, respectively	138
3.3.6 Conclusion	140
 Chapter 4: Insights into the pan-genomic binding site plasticity of HIF-1 and HIF-2, and the functional consequences of shared HIF-α binding sites....	
4.1 Introduction	142
4.2 Results	145

4.2.1 Validation of CRISPR KO cell line integrity	145
4.2.2 HIF binding site distribution and initial assessment of HIF- α binding site plasticity	149
4.2.3 Differential binding analysis of HIF ChIP-seq in HKC8 HIF- α knockout cells.....	153
4.2.4 Functional characterisation of shared HIF-1 and HIF-2 sites	159
4.3 Discussion	168
4.3.1 HIF does not undertake widespread redistribution to compensate for the absence of the other HIF- α isoform	169
4.3.2 The functional consequences of HIF-1 and HIF-2 shared binding sites	171
 Chapter 5: Cooperativity between HIF and AP-1	 173
5.1 Introduction	174
5.2 Results	178
5.2.1 AP-1 intervention utilising the AP-1 dominant-negative, AFos	178
5.2.2 Pan-genomic validation of AFos	184
5.2.3 AP-1, its pan-genomic association with HIF, and the effects of AFos on HIF DNA-binding.....	187
5.2.4 Differential binding of HIF in response to AFos	194
5.2.5 Characterisation of the pan-genomic distribution of HIF binding sites with respect to the nearest transcriptional start site.....	198
5.2.6 Attributing biological processes to HIF and AP-1 cooperativity	202
5.2.7 Association of HIF and AP-1 binding sites with gene expression	210
5.3 Discussion	216
5.3.1 Ablation of AP-1 DNA binding perturbs the pan-genomic binding of HIF	216
5.3.2 Cooperative binding of HIF and AP-1, and its potential relation to cancer	219
 Chapter 6: Final Discussion	 221
6.1 Canonical vs. non-canonical binding of HIF	222
6.2 Cell type specificity of HIF binding sites and binding site distributions ..	225

6.3 Inactivation of HIF- α and the determination of HIF compensatory capacity	231
6.4 Analysis of transcription factor binding in the vicinity of HIF binding sites	233
6.5 The effects of AP-1 intervention on HIF binding and function	235
6.6 Conclusion	237
Chapter 7: Materials and methods.....	238
7.1 Cell lines	239
7.2 Cell culture.....	239
7.3 Generic DNA techniques	239
7.3.1 Quantification of DNA concentration.....	239
7.3.2 Polymerase chain reaction (PCR) amplification of DNA.....	240
7.3.3 Restriction digest of DNA.....	240
7.3.4 Gel electrophoresis.....	241
7.3.5 Purification of amplified or digested DNA	241
7.3.6 Ligation of DNA fragments.....	241
7.3.7 Generation of new vector constructs	242
7.3.8 Bacterial strains, transformation and isolation of constructs.....	242
7.3.9 DNA and RNA sequencing	244
7.3.10 Transfection of plasmid DNA and lentivirus production	244
7.3.11 Lentiviral transduction using plasmid DNA	245
7.4 Protein techniques.....	245
7.4.1 Isolation of whole cell extracts	245
7.4.2 Immunoblotting	246
7.5 RNA techniques	247
7.5.1 RNA isolation	247
7.5.2 RNA quality control	247
7.6 Chromatin immunoprecipitation (ChIP)	248
7.6.1 Procedure for isolating ChIP DNA	248
7.6.2 DNA quality control	251
7.6.3 Quantitative analysis of ChIP DNA using SYBR green assay	252
7.7 Bioinformatics tools	252

7.7.1 ChIP-seq pipeline	252
7.7.2 BEDTools.....	253
7.7.3 SAMtools	253
7.7.4 Motif analysis and MEME-suite	254
7.7.5 Heatmaps and ngs.plot.....	254
7.7.6 Gene set enrichment analysis (GSEA)	254
7.7.7 Differential binding analysis	255
7.7.8 RNA-seq pipeline.....	255
7.8 Antibodies.....	256
7.9 Oligonucleotides	256
7.10 Solutions and buffers	257
7.10.1 Urea-Tris cell lysis buffer	257
7.10.2 Whole cell extract sample buffer (2x)	257
7.10.3 Western blot running buffer (10x)	257
7.10.4 Western blot transfer buffer (20x)	257
7.10.5 SDS-PAGE resolving gel buffer	258
7.10.6 SDS-PAGE stacking gel buffer	258
7.10.7 Phosphate buffered saline (PBS)	258
7.10.8 PBS-T	258
7.10.8 ChIP SDS lysis buffer	259
7.10.10 ChIP dilution buffer	259
7.10.11 ChIP low salt buffer.....	259
7.10.12 ChIP high salt buffer	260
7.10.13 ChIP lithium chloride buffer.....	260
7.10.14 TE buffer	260
7.10.15 ChIP elution buffer	261
Chapter 8: Appendix	262
8.1 Accession numbers for published NGS raw data.....	262
8.2 Accession numbers for ENCODE .bed files	263
8.3 Non-canonical HIF- α peak coordinates	267
8.4 ChIP-seq mapped read counts	268

8.5 AME analysis and significantly enriched motifs within specified sets of genomic regions	271
8.5.1 Sites preferentially bound by HIF-1 in HepG2	271
8.5.2 Sites preferentially bound by HIF-2 in HepG2	273
8.5.3 Sites preferentially bound by HIF-1 in RCC4	274
8.5.4 Sites preferentially bound by HIF-2 in RCC4	275
8.5.5 Sites preferentially bound by HIF-1 in HKC8	276
8.5.6 Sites preferentially bound by HIF-2 in HKC8	277
 Chapter 9: Bibliography	 279
 Chapter 10: Declaration	 293
 Chapter 10: Acknowledgements	 294

Chapter 1: Introduction

1.1 The hypoxia inducible transcription factor pathway

1.1.1 Oxygen, and the evolution of multicellular organisms

Oxygen has been fundamental to the evolutionary progression of eukaryotes and specifically metazoans by facilitating the highly efficient mitochondrial process termed oxidative phosphorylation. This process transfers chemical energy stored in carbon-carbon bonds, within organic entities such as glucose, to the high energy phosphate bond in adenosine triphosphate (ATP), which is subsequently used to power living cells (Semenza, 2012). Oxygen is essential for this process as it is required to act as the final electron acceptor in the electron transport chain. The energy production capacity enabled by mitochondrial aerobic respiration is now deemed to have been critical in producing sufficient energy to fuel the evolutionary complexity of multicellular organisms, which would not have been feasible through the use of glycolysis alone (Lane and Martin, 2010).

Primitive metazoans such as *Caenorhabditis elegans* depend upon the simple diffusion of oxygen from the atmosphere directly to all of the organism's cells, which consequently places restraints on the size of the organism. To circumvent these constraints, more complex organisms such as *Drosophila melanogaster* developed simple physiological systems to increase the surface area for diffusion and to reduce the diffusion distance thereby delivering sufficient oxygen to enable their bodies to expand to hundreds of thousands of cells. *Homo sapiens*, and other large vertebrates, have evolved even more complex networks of respiratory, circulatory and nervous systems in order to distribute oxygen to millions of cells, and maintain oxygen homeostasis

Within the mitochondria, electrons are transferred along the electron transport chain from electron donors such as NADH that are generated during

the citric acid cycle. This process creates a proton gradient across the mitochondrial membrane, that is subsequently used by ATP synthase to power the catalysis of ADP to high-energy ATP. Oxygen is the final electron acceptor in the oxidative phosphorylation electron transport chain and acute alterations in oxygen level can lead to aberrant electron transport chain activity. This may allow electrons to escape the electron transport chain and prematurely combine with oxygen, thus forming potentially harmful ROS (reactive oxygen species). Therefore, the accurate balancing of oxygen supply and utilisation is of critical importance to both the cell and the organism.

I have detailed the critical nature of oxygen for energy production in eukaryotes. I will now describe in detail the underlying mechanisms of cellular oxygen sensing and hypoxia inducible transcription factors (HIF), subsequently outlining how HIF acts as the primary regulator of oxygen homeostasis in metazoans.

1.1.2 Hypoxia inducible transcription factors and the cellular oxygen-sensing pathway

Hypoxia describes conditions in which local oxygen tension is decreased. This may result from increased oxygen consumption such as during growth or exercise, or during circumstances where oxygen availability is low such as at high altitude. Furthermore, important pathological conditions that affect gas transport in the lungs or disrupt local oxygen delivery by reducing blood supply as well as conditions of disordered growth such as cancer can also lead to hypoxia. Metazoans are capable of reacting to hypoxia by increasing the delivery of oxygen both systemically and locally or by reducing oxygen consumption within

cells. Rapid systemic adaptations such as ventilatory or cardiovascular responses are largely controlled by neuronal signalling, whilst more delayed cellular and endocrine/paracrine responses are largely under transcriptional regulation. Good examples of the transcriptional regulation of endocrine and paracrine factors that increase oxygen delivery to the tissues are erythropoietin (EPO) and vascular endothelial growth factor (VEGF), respectively. Conversely, transcriptional induction of anaerobic glycolysis and pathways that divert pyruvate to lactate rather than acetyl-CoA (the substrate for oxidative phosphorylation) are classic examples of reduced cellular oxygen demand in response to hypoxia, whilst maintaining ATP production.

The hypoxia-inducible factor (HIF) family of transcription factors are central to this cellular response to hypoxia (Kaelin and Ratcliffe, 2008; Semenza, 2012). HIF was first identified as a nuclear factor that bound to a DNaseI-hypersensitive and hypoxia-inducible enhancer region 3' to the human erythropoietin (*EPO*) gene in human hepatoma cells (Semenza et al., 1991; Wang and Semenza, 1993). The same group later purified and characterised the protein showing that it was composed of a heterodimer of HIF-1 α and HIF-1 β (aryl hydrocarbon receptor nuclear translocator, *ARNT*) subunits, both of which belong to the bHLH-PAS (basic helix-loop-helix PER-ARNT-SIM) superfamily of transcription factors (Wang and Semenza, 1995; Wang et al., 1995; Jiang et al., 1996). The HIF-1 α subunit is regulated by oxygen and stabilised in hypoxia, whereas levels of the beta subunit are unaffected by oxygen. Following stabilisation, in the classical model HIF-1 α heterodimerises with HIF-1 β , translocates to the nucleus and binds DNA. It should be noted that HIF- α can exist as several isoforms, and the similarities and distinctions between these

isoforms will be discussed during a later section of this Introduction.

Importantly, further studies on HIF-1 α demonstrated that it was expressed in virtually all cell types, irrespective of their ability to express *EPO*, suggesting that it had broader functions beyond the regulation of *EPO* (Maxwell et al., 1993). This was soon confirmed by the identification of additional targets such as *PGK1* (phosphoglycerate kinase 1) and *LDHA* (lactate dehydrogenase A) that are also regulated by HIF (Firth et al., 1994; Semenza et al., 1994).

A second HIF- α isoform, HIF-2 α (*EPAS1*), was subsequently identified by multiple groups (Ema et al., 1997; Flamme et al., 1997; Tian et al., 1997). Unlike HIF-1 α , which is ubiquitously expressed, the distribution of HIF-2 α is more restricted between different cell types, suggesting a more specialised function (Wiesener et al., 2003). A third HIF- α isoform, HIF-3 α (*HIF3A*), has also been identified based upon sequence homology to the other HIF- α isoforms, in addition to hypoxia inducibility (Gu et al., 1998). HIF-3 α exists in many different splice variants, and interestingly, one such variant has been demonstrated to act as a dominant-negative by forming inactive dimers with HIF-1 α (Maynard et al., 2003; Maynard et al., 2005; Tanaka et al., 2009). However, the focus of this thesis will be the two HIF- α isoforms with greatest similarity, HIF-1 α and HIF-2 α .

Both HIF-1 α and HIF-2 α dimerise with HIF-1 β through protein-protein interactions between their respective PAS domains to form intact HIF-1 and HIF-2 heterodimers. HIF-1 β is present in cells in excess of HIF- α , therefore HIF- α protein levels determine HIF transcriptional activity (Semenza et al., 1996). Notably, HIF-1 β is capable of dimerising with other bHLH-PAS proteins including the dioxin-inducible AhR (Lees and Whitelaw, 1999). HIF-1 α and HIF-2 α have both been shown to dimerise with ARNTL (Hogenesch., et al 1998) and ARNT2

(Ravasi., et al 2010; Corominas., et al 2014), however the functionality of these alternate HIF- α dimers is as yet unclear.

HIF-1 and HIF-2 interact with DNA via their bHLH DNA-binding domains. The DNA binding domains of both HIF-1 α and HIF-2 α are located within their N-terminus and share 83% amino acid sequence identity (Ema et al., 1997). The remarkable sequence similarity between HIF-1 α and HIF-2 α DNA binding domains, coupled to the shared dimerisation partner, leads to both factors binding the identical RCGTG consensus motif within DNA (Wenger et al., 2005; Schodel et al., 2011). This consensus motif is termed the 'hypoxia response element' (HRE). This core DNA recognition motif is non-palindromic, consistent with the heterodimeric nature of the intact HIF protein.

Oxygen sensing by the HIF- α proteins is conferred through two oxygen-dependent degradation domains (ODDD) and two transactivating domains (TADs) situated in the C-terminal part of the protein. These are termed the N-terminal ODDD (NODDD), C-terminal ODDD (CODDD), N-terminal TAD (N-TAD) and C-terminal TAD (C-TAD); with the CODDD and N-TAD partially overlapping each other. The 2-oxoglutarate (2-OG) dependent prolyl hydroxylase enzymes (PHDs) catalyse hydroxylation of one or both of two highly conserved proline residues within the NODDD and CODDD/C-TAD domains when oxygen is abundant (Epstein et al., 2001; Bruick and McKnight, 2001). In human HIF-1 α , these PHD enzymes hydroxylate Pro-402 and Pro-564 for hydroxylation within the NODDD and CODDDs, respectively. Hydroxylation of the proline residues permits interaction with the E3 ubiquitin ligase, von Hippel-Lindau (pVHL), leading to rapid proteasomal degradation of HIF- α in the presence of oxygen (Maxwell et al., 1999). A schematic highlighting relative localisation of the

conserved domains between the HIF isoforms can be identified in Figure 1.

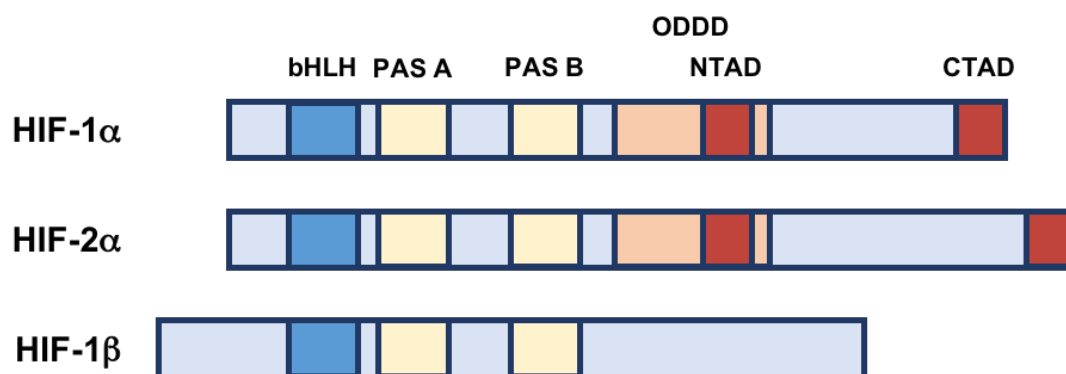


Figure 1. Schematic of hypoxia inducible factor domain structure.

bHLH: basic helix-loop-helix; PAS A/B: per-ARNT-sim domain A or B; ODDD: oxygen-dependent degradation domain; NTAD: N-terminal transactivation domain; CTAD: C-terminal transactivation.

Highlighting the importance of HIF and the PHD/VHL oxygen-sensing pathway is the high degree of conservation in metazoans. The system is conserved in the most primitive animal, *Trichoplax adhaerens*, as well as all other extant metazoans tested to date (Loenarz et al., 2011). Gene duplication events during evolution have meant that higher order metazoans express three different PHD enzymes, termed PHD1-3 (*EGLN2*, *EGLN1*, *EGLN3*), whereas lower order organisms such as *C. elegans* only express a single isoform (EGL-9). Of the three human PHD isoforms, PHD2 is the most related to the ancestral isoform by sequence identity (Rytkönen et al., 2011), particularly with respect to the conservation of the PHD MYND finger domain (Loenarz et al., 2011).

A third HIF residue, within the C-TAD (Asn-803 in HIF-1α), is targeted for oxygen-dependent hydroxylation by a different 2-oxoglutarate-dependent dioxygenase, known as Factor Inhibiting HIF-1 (FIH). Hydroxylation of the asparagine residue within the CTAD of HIF-α, inhibits interaction with the transcriptional co-activator p300, thereby repressing the transactivating activity of HIF in the presence of oxygen (Mahon et al., 2001; Lando et al., 2002). However,

whilst FIH is inhibited by hypoxia in the same way as the PHD enzymes, its inhibition requires more severe hypoxia than is required to inhibit the PHD enzymes (Tian et al., 2011).

The mechanism by which the PHDs and FIH hydroxylate their target prolyl and asparagine residues is via the splitting of O₂, whereby one oxygen atom is used in the oxidative decarboxylation of 2-oxoglutarate yielding succinate and CO₂, whilst the other oxygen atom is incorporated into the oxidised target amino acid side chain within HIF- α (Hewitson et al., 2002; McNeill et al., 2002).

1.2 HIF-1 and HIF-2 function *in vivo*

1.2.1 The role of HIF in development

Mammalian embryogenesis is understood to be a highly hypoxic process (Dunwoodie, 2009), therefore it is perhaps unsurprising that both HIF-1 α and HIF-2 α are essential for mammalian development. Mice homozygous for a “null” HIF-1 α allele die by embryonic day 10.5 owing to vascular abnormalities, cardiac defects and defective erythropoiesis (Iyer et al., 1998). Similarly, HIF-2 α “null” mice die at embryonic day 12.5 owing to vascular defects (Peng et al., 2000). However, depending on the genetic background of the mice being studied, the cause of death in HIF-2 α null knockout mice can vary. For instance, mice can die due to impaired catecholamine production leading to lethal bradycardia during embryogenesis (Tian et al., 1998), or at a neonatal stage due to defective lung maturation (Compernelle et al., 2002).

Conditional HIF- α knockout studies have illustrated tissue-specific functions for HIF during development. Whilst there is a broad array of elegant studies, I will briefly describe some examples. HIF-1 α was deleted in the

cartilaginous growth plate of developing bone, an avascular and thus hypoxic tissue, subsequently highlighting the requirement for HIF-1 α to mediate growth arrest during chondrogenesis (Schipani et al., 2001). Adipocyte differentiation is augmented by HIF-1 α deficiency in mouse embryonic fibroblasts, thus leading to PPAR gamma-2 activation and excessive adipogenesis (Yun et al., 2002). Mice lacking HIF-1 α in osteoblasts developed abnormally thin and under-vascularised bones, thus demonstrating a requirement for functional HIF-1 α to mediate proper osteogenesis (Wang et al., 2007). By deleting HIF-1 α in haematopoietic stem cells (HSCs), HSC responsiveness to stress triggers was abrogated due to the requirement of HIF-1 α for effective maintenance HSC cell cycle and stem cell quiescence within the endosteal of the bone marrow (Takubo et al., 2010).

1.2.2 Physiological and pathophysiological roles of HIF

As previously outlined, in hypoxia, cells shift their metabolic flux from oxidative phosphorylation to anaerobic glycolysis. This process is at least in part mediated by HIF, wherein HIF up-regulates glycolytic genes that potentiate glycolysis and reduce metabolite entry into the citric acid cycle (Semenza et al., 1996; Kim et al., 2006). Concurrently, HIF also up-regulates factors associated with reducing mitochondrial biogenesis (Zhang et al., 2007) and increasing mitochondrial autophagy (Zhang et al., 2008a), therefore down-regulating oxidative phosphorylation.

Studies have suggested that this metabolic reprogramming by HIF is cytoprotective during myocardial infarction (Shohet and Garcia, 2007). Intermittent cardiac ischaemia, evoked by intermittent coronary artery occlusion, has been shown to be a beneficial “pre-conditioning” strategy to protect against

longer ischaemic events or from myocardial infarcts in mouse models. HIF-1 α has been shown to be involved in this process, as mice deficient for HIF-1 α have a reduced capacity to prevent tissue loss during myocardial infarction and have a higher rate of mortality in these pre-clinical models (Shohet and Garcia, 2007; Cai et al., 2008).

Angiogenesis is required for cutaneous wound healing, whereby both systemic and local responses contribute towards the vascular remodeling process. Angiogenesis describes the process of the formation of new blood vessels from pre-existing vessels. The angiogenic process is initiated from within the injured tissue via the activation of endothelial cells. HIF-1 activation evokes the upregulation of cytokines that activate and recruit BMDACs (bone marrow derived angiogenic cells) which participate in the process of angiogenesis. Impairments in this process are strongly associated with the synergistic effects of aging and diabetes (Brem et al., 2007).

1.2.3 The role of HIF in cancer

As a consequence of rapid cell division in a primary tumour, the degree of hypoxia within a tumour mass can vary significantly. Whilst prolonged oxygen depletion (anoxia) will evoke necrosis, viable hypoxic tumour cells are known to surround necrotic regions of a tumour. Numerous studies have demonstrated that increased intratumoural hypoxia levels are associated with poor prognosis and resistance to radiation therapy in head and neck cancer, ovarian cancer, and oesophageal cancer (Zhong et al., 1999; Talks et al., 2000). In a wide array of studies, the hypoxia inducible factors have been shown to influence numerous aspects of tumourigenesis, including initiation, progression, invasion,

inflammatory cell recruitment, and metastasis.

HIF-2 α has been associated with cancer stem cell function and tumour initiation. HIF-2 α , and not HIF-1 α , is a regulator of the OCT4 (*POU5F1*) gene that is responsible for maintaining the undifferentiated state of embryonic stem cells (Covello et al., 2006). HIF-2 α has been associated with “stemness” via its selective expression in the CD133+ population of glioblastoma ‘cancer stem cells’, whereas HIF-1 α has been shown to be expressed in both the tumour and non-tumour populations of CD133+ cells (Li et al., 2009b). In human neuroblastoma, there is a stem cell sub-population that is defined by its expression of vimentin, ID2, NOTCH1, HES1, and also HIF-2 α (Pietras et al., 2009). Silencing of HIF-2 α in xenograft neuroblastoma mouse models resulted in a partially-necrotic and poorly vascularised tumour that expressed markers of sympathetic neuronal cells, whereas control tumours maintained stem cell markers and were well vascularised, thus suggesting that HIF-2 α is important for the regulation of tumour-initiating cancer stem cells.

Mouse tumour xenograft models incorporating a HIF-1 α or HIF-2 α intervention have highlighted some distinct roles of the HIF- α isoforms on tumour progression. HIF-1 α deficient teratoma xenografts have reduced growth and angiogenesis, therefore suggesting that HIF-2 α may be prohibitive for teratoma growth (Ryan et al., 1998; Carmeliet et al., 1998). Further corroborating this, HIF-2 α deficient teratoma xenografts have increased growth (Acker et al., 2005). HIF-1 α appears to have play an oncogenic role in fibrosarcoma wherein xenografts deficient for HIF-1 α exhibit a reduced capacity for growth (Ryan et al., 2000). Interestingly, in clear cell renal carcinoma xenografts, HIF-1 α and HIF-2 α appear

to have opposing function whereby overexpression of HIF-1 α reduced tumour growth, whilst HIF-2 α overexpression enhanced tumour growth (Raval et al., 2005). Conversely, in a KRAS-driven NSCLC (non-small cell lung carcinoma) autochthonous tumour model, conditional knockout of HIF-1 α had no effect on tumour burden, whereas conditional knockout of HIF-2 α enhanced tumour progression (Mazumdar et al., 2010).

The hypoxia associated factor (HAF) is an E3 ubiquitin ligase that has been shown to physically associate with HIF-1 α to contribute towards its proteasomal degradation during acute hypoxia (Koh et al., 2008). HAF also interacts with HIF-2 α but does not facilitate its proteasomal degradation, instead it increases the transactivating capacity of HIF-2 (Koh et al., 2011). Under HAF overexpression conditions, the balance of hypoxia-signaling shifts towards HIF-2 α , subsequently leading to the up-regulation of HIF-2 α target genes such as *POU5F1* and *MMP9*, consequently leading to the increased invasiveness of a number of cancer cell lines during *in vivo* invasion assays.

Tumour associated macrophages have been demonstrated to facilitate neoplastic transformation, immune cell evasion and metastasis (Williams et al., 2016). Interestingly, HIF-1 α and HIF-2 α appear to oppose one another in the regulation of nitric oxide (NO) in macrophages. NO is an important feature of tumour associated macrophages in that NO production inhibits T-cell activity and immune surveillance *in vivo* (Doedens et al., 2010). HIF-1 α deficient macrophages expressed reduced levels of inducible nitric oxide synthase (iNOS), whilst conversely, HIF-2 α up-regulates the expression of arginase-1 which reduces the availability of substrate for NO production. Therefore, HIF-2 α is considered to be prohibitive in the recruitment and activation of tumour

associated macrophages.

Finally, through the use of the MMTV-PyMT (mouse mammary tumour virus-polyoma middle T antigen) autochthonous tumour model, conditional knockout of HIF-1 α in mouse mammary epithelial cells was shown to elicit a significant reduction in pulmonary metastasis (Liao et al., 2007).

These examples of HIF- α activity in *in vivo* tumour models demonstrate context-specific roles of HIF-1 and HIF-2 that in certain circumstances may be antagonistic of one another.

1.3 Pan-genomic approaches to studying HIF

1.3.1 Regulation of gene expression by HIF

In the late 1990s and early 2000s, increasing numbers of studies were unveiling individual hypoxia-inducible genes that appeared to be regulated by HIF (reviewed in Semenza, 2003). An initial microarray-based assessment of the hypoxic transcriptome seemed to suggest that HIF-2 α was dispensable for hypoxic regulation (Park et al., 2003). Specifically, in MEFs (mouse embryonic fibroblasts), HIF-2 α was shown to be insensitive to oxygen tension whilst also being constitutively trapped in the cytosol. HIF-1 α “null” MEFs were no longer capable of up-regulating *Pgk*, *Glut-1* or *Vegf* in hypoxia, therefore hypoxia-inducible gene expression in MEFs was solely mediated via HIF-1. However, in preceding *in vivo* studies, HIF-2 α has been demonstrated to play an important and varied role mouse embryogenesis, therefore further gene regulatory studies were consequently performed to better understand the respective roles of HIF-1 α and HIF-2 α in hypoxic gene regulation.

The Ratcliffe Group, amongst others, have shown that HIF-1 and HIF-2 are in fact capable of regulating both unique and common target genes *in vivo* and in multiple cell lines (Sowter et al., 2003; Hu et al., 2003; Grabmaier et al., 2004; Warnecke et al., 2004; Rankin et al., 2005; Raval et al., 2005; Wang et al., 2005; Covello et al., 2006; Gruber et al., 2007). For example, HIF-1 specifically regulates glycolytic genes (including phosphoglycerate kinase [*PGK*] and lactate dehydrogenase A [*LDHA*]) (Hu et al., 2003; Rankin et al., 2005; Wang et al., 2005), as well as carbonic dehydratase-9 (*CA9*) (Grabmaier et al., 2004) and *BNIP3* (Raval et al., 2005), whereas HIF-2 exclusively regulates the Pou transcription factor Oct-4 (*POU5F1*), cyclin D1 (*CCND1*), and transforming growth factor- α (*TGFA*) (Raval et al., 2005; Covello et al., 2006). Other hypoxia-inducible genes, such as vascular endothelial growth factor (*VEGF*), facilitated glucose transporter-1 (*SLC2A1*), adipose differentiation-related protein (*ADRP*), adrenomedullin (*ADM*), and N-myc downstream regulated 1 (*NDRG1*) are regulated by both HIF-1 and HIF-2 (Hu et al., 2003; Raval et al., 2005; Hu et al., 2006). These results are in agreement with functional studies indicating that HIF-1 and HIF-2 exhibit distinct roles during both development and tumor progression (Maranchie et al., 2002; Kondo et al., 2003; Scortegagna et al., 2003; Covello et al., 2005, 2006; Raval et al., 2005).

A recent study has shed new light on the non-coding RNA transcriptome regulated by hypoxia (Choudhry et al., 2014). These RNA-seq analyses utilised total RNA rather than polyA⁺ selected mRNA, thus leading to the revelation that all classes of RNA are profoundly regulated by hypoxia. The functional consequences of the hypoxia-induced non-coding transcriptome remain largely unexplored.

Understanding what hypoxia-regulated genes are direct HIF targets vs. those that are secondary effects is not possible from transcriptomics analysis alone. An understanding is therefore required of the sites that HIF-1 and HIF-2 actually bind within the genome. Although, in circumstances in which the distance between a HIF binding site and the nearest gene is in the tens or hundreds of kilobases, chromatin conformation capture technology may be required to determine which gene within a topologically associated domain (TAD) is the direct HIF target (To be discussed further in section 1.3.3).

1.3.2 Pan-genomic HIF DNA binding studies

In the light of the increasing numbers of hypoxic transcriptome studies being published in the mid to late 2000s, efforts were made to understand which hypoxia-regulated genes were direct HIF targets. Prior to the adequate development of chromatin immunoprecipitation (ChIP) and next-generation sequencing (NGS) technologies, *in silico* analysis utilised transcription factor binding motif position weight matrices to predict binding sites via coupling to high confidence hypoxia regulated genes (Ortiz-Barahona et al., 2010). This study predicted 216 HIF target loci, several of which were successfully validated by ChIP-qPCR. However, this type of methodology (used by other groups in alternate contexts; see Bussemaker et al., 2001; Jeffery et al., 2007; Hoek et al., 2008) is low throughput and potentially biased towards the identification of binding sites proximal to regulated genes.

ChIP coupled to pan-genomic methodologies such as tiled micro-arrays (ChIP-chip) and next generation sequencing (ChIP-seq) have provided a high-throughput approach to identify protein-DNA binding events genome-wide (Park,

2009). Early attempts using ChIP-chip (Mole et al., 2009; Xia et al., 2009) led to the identification of several hundreds of HIF binding sites in MCF7 breast cancer cells and HepG2 hepatoma but have now been largely superseded by ChIP-seq to identify HIF-1 α , HIF-2 α and HIF-1 β binding (Schodel et al., 2011; Villar et al., 2012; Schodel et al., 2012; Choudhry et al., 2014; Tausendschön et al., 2015; Salama et al., 2015; Grampp et al., 2016; Platt et al., 2016).

Although ChIP-chip and ChIP-seq are able to identify HIF-binding sites genome-wide, this does not necessarily identify the transcriptional target of each site. Schodel et al coupled ChIP-seq data with gene regulation data from RNA-seq analyses of normoxic and hypoxic cells (Schodel et al., 2011). In this analysis HIF binding sites are associated with genes that are upregulated in hypoxia, but not with genes that are down-regulated, indicating that HIF acts in *cis* as a transcriptional activator but not as a repressor. To a first approximation, the regulated gene is most often the most proximal gene to the HIF-binding site. However, at HIF-binding sites where the nearest gene is not regulated, the next nearest gene is likely to be induced indicating that HIF binding does not always regulate the nearest gene.

Schodel et al also identified several other characteristics of HIF binding sites. Firstly, HIF binding sites lay almost exclusively within regions of chromatin that were accessible (as defined by DNase-seq analysis) and already active as promoters or enhancers (as demonstrated by histone H3K4me1, H3K4me3 and H3K27ac ChIP-seq signal) in normoxia (i.e. in the absence of HIF). This suggests that HIF does not act as its own “pioneer” factor, but rather that it binds to enhancers or promoters that have been defined by other factors. Importantly, only a small proportion of accessible HRE motifs were occupied by HIF indicating

that other factors must help define HIF binding to chromatin. Secondly, motif analysis of HIF-binding regions confirmed the core consensus HRE motif (RCGTG) as the most enriched DNA motif at HIF-binding sites. Importantly, this motif was the same at HIF-1 and HIF-2 binding sites and further analysis of bases flanking this motif did not show any differences between HIF-1 and HIF-2 binding sites. However, despite this similarity in binding motif, HIF-1 and HIF-2 bound distinct and only partially overlapping sets of sites. In particular, HIF-1 bound more frequently closer to gene promoters, whilst HIF-2 displayed a more promoter-distal binding distribution. However, the generality of these observations, the mechanisms restricting binding of HIF to specific accessible HRE motifs, and the mechanisms underlying the differences in isoform binding remain unclear.

One possibility to explain each of these observations is the potential for cooperative interaction between HIF and other transcription factors that might help direct binding of one or both isoforms. In one analysis, the del Peso Group identified HIF-1 binding sites that were conserved between ChIP-chip datasets in at least two out of four different cell types (HeLa, MCF7, U87 and HepG2) (Villar et al., 2012). Motif analysis of binding regions identified several transcription factor motifs that were enriched relative to a set of background DNA sequences. Importantly, several stress responsive transcription factor motifs were enriched, including AP-1 (activator protein-1), CREB (cAMP response element-binding protein) and CEBP (CCAAT-enhancer binding protein). Moreover, experimental validation at HIF regulated promoters suggested that these factors may functionally influence HIF-mediated transactivation of specific target genes.

In a second study, the Ratcliffe Group performed ChIP-seq in 786O cells

expressing either exogenous HIF-1 α or HIF-2 α (Salama et al., 2015). Similarly to Villar et al, they observed enrichment of the AP-1 motif at HIF binding sites. However, they also observed a greater enrichment for the AP-1 motif at HIF-2 binding sites than at HIF-1 binding sites. Taken together, these studies suggest that there is a possibility of widespread cooperative interaction between the AP-1 and HIF transcription factors acting in *cis* at HIF binding sites.

1.3.3 HIF and chromatin conformation

It is well known that chromatin does not exist as a linear entity. Chromatin exists as an intricate three-dimensional network of DNA and associated proteins, whereby distal enhancers are capable of physically associating with their target gene via chromatin looping (Dekker et al., 2013). Prior to the development of chromosome conformation capture technologies, it was unclear what gene a promoter-distal intergenic binding site may be regulating.

A GWAS-identified polymorphism in ccRCC led to the identification of a previously unrecognised HIF-bound enhancer for cyclin D1 (*CCND1*), situated approximately 220 kb from the TSS (transcription start site) of *CCND1* (Schodel et al., 2012). The SNP is a protective polymorphism that impairs HIF binding. Chromosome conformation capture (3C) demonstrated ccRCC-specific looping from the HIF binding site to the TSS of *CCND1*. Interestingly, the physical interaction between the enhancer and *CCND1* was present in renal cells but not MCF7 breast cancer cells.

In a similar vein to the previously discussed GWAS and chromatin looping event, another ccRCC GWAS polymorphism was identified within a HIF binding site (Grampp et al., 2016). Chromatin conformation capture illustrated chromatin

looping from the HIF binding site to the two nearest genes, *PVT1* and *MYC*, however only enhanced *MYC* expression was associated with the GWAS risk allele in ccRCC patients.

Capture C has been utilised by the Ratcliffe Group in an effort to define direct gene targets associated with HIF binding sites, whilst also revealing how hypoxic induction of HIF influences 3D chromatin architecture connecting distant functional elements (Platt et al., 2016). Capture C is a technique that multiplexes 3C by oligonucleotide capture of regions of interest, and next generation sequencing (Hughes et al., 2014; Davies et al., 2015). Platt et al defined long distance enhancer-promoter interactions, thus confirming direct HIF target genes with respect to approximately 30 HIF binding sites. Importantly, this study also demonstrated that 3D chromatin architecture does not appear to change substantially in response to hypoxia or prolonged VHL inactivation, therefore HIF acts mainly upon pre-existing chromatin-chromatin interactions.

1.4 Transcription factor binding site selection

1.4.1 Transcription factor cooperativity

Transcription factors bind DNA at regions comprising their consensus motif, however the presence of a consensus motif alone is not sufficient for determining a transcription factor binding site. As an example, there are over 2×10^6 HREs within the genome, however HIF only binds to a subset of these, even within accessible regions of chromatin (Schodel et al., 2013). In the majority of instances, accessible chromatin is necessary for a transcription factor to bind DNA (Thurman et al., 2012), however even among accessible chromatin there remains a requirement to employ mechanisms of selectivity.

A number of studies have demonstrated cooperativity of transcription factors in both the regulation of gene expression, and in DNA binding site selection. Studies have demonstrated gene regulatory cooperativity between HIF and AP-1 with respect to a small number of genes, however there are yet to be any published studies illustrating genome-wide DNA binding cooperativity between HIF and another transcription factor. *VEGFA* and *EDN1* expression in hypoxia are postulated to be a result of probable DNA-binding cooperativity between HIF and AP-1 transcription factors (Damert et al., 1997; Yamashita et al., 2001). Luciferase reporter assays demonstrated that to achieve full hypoxic induction of *VEGFA* and *EDN1*, both the HRE within the promoter of each gene, along with the upstream AP-1 consensus motif, is required for maximal hypoxic induction.

Substantial insight has been gained into the biological activity of certain transcription factors following transcription factor pathway intervention and subsequent pan-genomic binding analysis. These methodologies have uncovered relationships between two or more transcription factors that influence the phenotypic output of a cell. Uncovering transcription factors that cooperate with HIF on the pan-genomic level remains a challenge for the field, however this thesis later delivers early insight into one such cooperative transcription factor relationship with HIF.

An early example of transcription factor cooperativity was illustrated by Jason Carroll's laboratory, one of the pioneering groups for ChIP-seq studies. The forkhead box (FOX) transcription factor motif was shown to be significantly enriched within estrogen receptor binding sites relative to random background genomic regions, and the FOXA1 transcription factor was shown to be required

for estrogen receptor (ER) to bind to the entirety of its target *cis*-regulatory elements in ER+ breast cancer cells (Carroll et al., 2005; Laganière et al., 2005). FOXA1 was later described as a pioneer factor, whereby the forkhead transcription factor maintained open chromatin for ER at its target *cis*-regulatory elements prior to induction by estrogen (Zaret and Carroll, 2011).

AP-1 has been shown to maintain chromatin accessibility, thus enabling the glucocorticoid receptor (GR) to bind approximately half of its target *cis*-regulatory elements in mammary epithelial cells (Biddie et al., 2011). A similar mechanism is mediated by the C/EBP in hepatic cells in order to facilitate the binding of the GR to a significant proportion of its target loci (Grontved et al., 2013).

More recently, AP-1 has been shown to cooperate with the TEAD family of transcription factors, whereby AP-1 facilitates the binding of TEAD to active *cis*-regulatory elements (Zanconato et al., 2015), whilst cooperating in the activation of a transcriptional programme to evoke cell migration and invasion (Liu et al., 2016).

In each of the described pan-genomic cooperativity studies, it might be argued that the investigators have been fortuitous in that they have identified a cooperating transcription factor on the foundation of identifying a transcription factor binding motif within the binding sites of the primary transcription factor of interest. This may be considered fortuitous because interpreting motif enrichment has a number of caveats. Firstly, as previously stated with respect to the HRE, the presence of a motif does not necessarily mean that the transcription factor will be bound. Secondly, multiple transcription factors may be capable of binding the same or a closely related motif, so it may not be possible to know which of

these transcription factors is actually binding. Thirdly, an extension of this is that transcription factors that bind to motifs that share homology with the consensus motif of interest (e.g. an E-box, CACGTG, contains a HRE, RCGTG) may be enriched as a consequence of enriching for the primary motif and therefore not representing co-incident binding at all. Finally, GC content varies throughout the genome and is generally higher close to gene promoters, therefore a GC-rich motif is inherently more likely to be enriched within a set of binding sites that are promoter-proximal.

To circumvent the previously-described motif analysis pitfalls, a technique has been developed by Jason Carroll's laboratory to determine the chromatin-bound interactome of transcription factors of interest. This technique is termed RIME (rapid immunoprecipitation mass spectrometry of endogenous proteins) and was used during the identification of GREB1 as an ER cooperating factor (Mohammed et al., 2013). The technique involves a broadly similar procedure to that of ChIP, as far as the formaldehyde cross-linking and immunoprecipitation is concerned, however rather than the DNA being harvested from the chromatin cross-linked fraction of cell material, the immunoprecipitated cross-linked protein is digested and subjected to mass spectrometry. It is therefore possible to identify proteins in complex with the bait protein following mass spectrometry. GREB1 was subsequently shown to facilitate the interaction between ER and additional co-factors, thus enabling transcriptional activity of ER. A further study has successfully utilised this technology to identify an interaction between FOXK2 and BAP1, and demonstrated the necessity for this interaction in order to enable BAP1 recruitment to a subset of its anti-tumourigenic target genes (Ji et al., 2014). RIME could feasibly be applied to HIF in order to identify isoform-specific

or cell type-specific regulatory factors that influence binding site selectivity.

1.4.2 Hypoxia, histone modifications and DNA methylation

Covalent histone modification and DNA methylation are both examples of epigenetic traits. An early example of how an epigenetic trait influences the HIF response was illustrated by a study showing that HIF-1 requires the CpG within the HRE of the 3' *EPO* enhancer to be methylation-free if HIF is to be capable of binding said HRE (Wenger et al., 1998). However, the overall relationship between epigenetics, hypoxia and the HIF pathway remains largely unexplored.

Hypoxia has recently been shown to evoke DNA hypermethylation as a consequence of reducing the activity of the TET methylcytosine dioxygenases (Thienpont et al., 2016). The study emphasised how the increase in cytosine methylation was independent of HIF, however it failed to demonstrate how the increased CpG methylation affects pan genomic binding of HIF. It therefore remains unclear to what extent DNA methylation influences HIF DNA binding across the genome.

Early HIF ChIP-chip and ChIP-seq studies demonstrated an association between binding sites and the levels of certain histone modifications. HIF-1 binding is associated high levels of H3K4me3 (Xia et al., 2009), a mark of active promoters. In contrast, H3K27me3, a mark of inactive enhancers, is absent from HIF-1 binding sites (Tanimoto et al., 2010). H3K27ac has been shown to increase in hypoxia at the promoter of *CA9* (Ledaki et al., 2015), and to be associated with an increase in chromatin accessibility at the same locus. Moreover, it has recently been demonstrated that a significant subset of HIF-bound enhancers and target gene promoters exhibit increased levels of H3K27ac

in hypoxia (Platt et al., 2016). However, it is unclear whether these changes in histone modification bear any mechanistic relationship to HIF binding.

It is possible that HIF-1 and HIF-2 both adopt their respective promoter-proximal and promoter-distal binding sites due to an affinity for a particular histone environment, or via associations with histone modification readers, thus underpinning the differential binding site selectivity between the two HIF- α isoforms.

1.5 Aims of this study

The importance of the HIF pathway in the physiological and pathophysiological response to hypoxia is well documented. To date, little is known about the determinants of HIF-1 and HIF-2 binding site selection, aside from the proposed requirement for HIF- α to heterodimerise with HIF-1 β , and the binding of the consensus HRE within DNA. Questions remain over why and how HIF-1 and HIF-2 bind both shared and distinct sites, whilst regulating overlapping and distinct sets of genes. In this thesis, I aim to identify HIF-1 and HIF-2 binding sites across a selection of cell lines using high quality ChIP-seq data, whilst identifying characteristics of HIF-1 and HIF-2 binding sites which may enable greater understanding of the mechanisms that underpin HIF binding site selection.

HIF- α has previously been demonstrated to function transcriptionally in the absence of HIF-1 β (Gustafsson et al., 2005; Xiang et al., 2015; Salama et al., 2015). The first part of this thesis aims to determine the extent to which genome-wide HIF- α DNA binding is canonical, i.e. in heterodimeric complex with HIF-1 β (Chapter 2).

HIF-1 and HIF-2 binding sites have previously been postulated to be functionally associated with other transcription factors via motif analysis (Villar et al., 2012; Salama et al., 2015). Furthermore, HIF-1 and HIF-2 binding sites have been shown to differ with respect to their propensity to bind close to or distant from the nearest gene (Schodel et al., 2011). The second part of this thesis describes analysis to determine whether the binding distributions of HIF-1 and HIF-2 are conserved across cell types, whilst also performing several analyses in order to draw more significant associations between HIF binding sites and other transcription factors, as well associations with the local histone environment (Chapter 3). The associations established provided the subject for the fourth section of this thesis (Chapter 5), whereby an intervention has been made against AP-1 in order to establish whether AP-1 is contributing towards pan-genomic HIF binding site selection and HIF-mediated gene regulation.

Another possible factor influencing HIF binding site selection is that preferential binding of one HIF-isoform (by whatever mechanisms) blocks binding of the other isoform, in which case overexpression and/or depletion of one isoform might lead to a compensatory re-distribution of isoform specific binding. Expression of exogenous HIF-1 α or HIF-2 α at supra-physiological levels in 786O cells did not lead to a major re-distribution of binding. Specifically, the promoter-proximal versus promoter-distal distribution of HIF-1 and HIF-2 binding respectively was preserved even at very high levels of each isoform. However, effects may be saturated at physiological levels of HIF- α subunits and I have therefore explored this possibility further in Chapter 4 of my thesis.

Specifically, the experiments outlined in this thesis were performed to provide:

- a) an understanding of whether HIF- α has the capacity to bind at sites in the absence of HIF-1 β ,
- b) a comparison of the degree of binding site conservation exhibited by HIF-1 and HIF-2 across four cell lines,
- c) an establishment of whether the binding site distributions of HIF-1 and HIF-2, with respect to distance between a site and the nearest gene, is conserved across cell types,
- d) an indication of what transcription factors are preferentially associating with HIF-1 or HIF-2 binding sites,
- e) an indication of what histone environments are preferentially targeted by HIF-1 or HIF-2,
- f) an insight into the capacity of HIF-1 and HIF-2 to exhibit binding site plasticity and subsequently redistribute in the absence of the other HIF- α isoform,
- g) an insight into the conditions that enable prediction of when the nearest gene to a shared HIF-1 and HIF-2 binding site is to be exclusively regulated by either HIF-1 or HIF-2,
- h) an indication of the extent to which HIF recruitment to its target loci, and subsequent HIF-mediated gene regulation, is influenced by AP-1 on a pan-genomic scale in renal cancer.

Chapter 2: Understanding the concordance of HIF- α / β binding to chromatin, and identification of any instances of non-canonical HIF- α binding

2.1 Introduction

Hypoxia-inducible factor 1 (HIF-1) consists of the subunits HIF-1 α and HIF-1 β . Both HIF-1 α and HIF-1 β contain basic helix-loop-helix-PAS (bHLH-PAS) domains (Wang et al., 1995) that enable HIF- α/β heterodimerisation and subsequent DNA binding (Jiang et al., 1996). HIF-2 α is a paralog of HIF-1 α , that likewise contains bHLH and PAS domains and heterodimerises with HIF-1 β to form the intact HIF-2 heterodimer (Tian et al., 1997; Ema et al., 1997; Hogenesch et al., 1997). Following stabilisation of HIF- α subunits in hypoxia, both HIF-1 and HIF-2 heterodimers, bind hypoxia response elements (HREs) within the genome (Wenger et al., 2005), and transactivate target gene expression (Schodel et al., 2011).

Next-generation sequencing (NGS) coupled to chromatin immunoprecipitation (ChIP-seq) has enabled the identification of transcription factor binding sites, as well loci harbouring specific histone modifications, across the genome (Park, 2009). Several studies have performed pan-genomic HIF binding assays (Mole et al., 2009; Xia and Kung, 2009; Tanimoto et al., 2010; Schodel et al., 2011; Villar et al., 2012; Schodel et al., 2012; Choudhry et al., 2014; Tausendschön et al., 2015; Salama et al., 2015; Grampp et al., 2016; Platt et al., 2016), although the majority of the early studies used highly stringent criteria to define HIF-1 and HIF-2 binding sites from ChIP-seq data. For instance, initial ChIP-on-chip data by Xia et al., required the presence of a HRE motif (RCGTG) to define a HIF-1 binding site in HepG2 cells. Meanwhile, Schödel et al., used overlap with HIF-1 β ChIP-seq data and concordance with prior ChIP-on-chip data (Mole et al., 2009) to define HIF-1 α and HIF-2 α binding sites in MCF7 cells. These initial pan-genomic HIF binding studies were performed during the

early phases of NGS technology development, and may have underestimated the extent of HIF binding. Specifically, the requirement that HIF binding sites must contain a HRE motif or that HIF- α and HIF-1 β must co-bind, ignores the possibility that HIF- α subunits may be able to bind HRE- or non-HRE-containing sequences on chromatin without HIF-1 β . This could potentially occur through indirect tethering of HIF to DNA via interaction with a pre-formed chromatin-bound transcription factor complex or direct binding of HIF- α subunits to chromatin either as monomers, homodimers or as heterodimers with other (bHLH-PAS) transcription factors. Indeed, previous pan-genomic studies have outlined functional transcription factor tethering, and widespread instances of transcription factors binding to atypical motifs, both in the context of estrogen receptor alpha (ER) and activator protein-1 (AP-1) (Gertz et al., 2013; Zanconato et al., 2015).

A similar mechanism of non-canonical chromatin interaction (HIF- α binding to chromatin in the absence of HIF-1 β) has been reported for HIF-1 α which was shown to bind a HRE-less site within the promoter of *CTGF* through a tethering interaction with the transcription factor, TAZ (Xiang et al., 2015). A prior study has also demonstrated a tethering function of HIF-1 α . In the absence of HIF-1 β , HIF-1 α binds the NOTCH intracellular domain and is subsequently recruited to NOTCH-responsive promoters such as *HES* and *HEY* (Gustafsson et al., 2005).

In a recent pan-genomic HIF binding study, published by the Ratcliffe group, thousands of canonical *and* non-canonical HIF- α binding sites were identified (Salama et al., 2015). The non-canonical sites were considered genuine given a functional association with induced expression of the

neighbouring gene in response to overexpression of either HIF-1 α or HIF-2 α . However, in this study, the pan-genomic binding of overexpressed HIF-1 α and HIF-2 α rather than endogenous proteins were assayed, and this may have resulted in non-physiological binding.

Therefore, I wished to establish whether endogenous HIF-1 α and HIF-2 α are capable of exhibiting widespread non-canonical binding. To aid this investigation, I generated new, high depth ChIP-seq data across three separate cell lines in biological duplicate. I subsequently subjected these data to various methods of analysis in order to fully determine the overall levels of concordance between HIF- α and HIF-1 β . Specifically, these methods include:

1. Venn diagrams to illustrate extent of binding site overlap between sets of MACS called peaks.
2. ChIP-seq signal correlative analyses between HIF- α and HIF-1 β within sites called by MACS as overlapping or non-overlapping.
3. Heatmap analysis to indicate whether HIF- α peaks called by MACS ever exhibit no greater HIF-1 β ChIP-seq signal intensity than the local HIF-1 β background signal.
4. A further quantitative analysis that incorporates a genome-wide HIF-1 β background signal capable of corroborating the heatmap analyses, whilst also enabling easier identification of isolated instances of HIF- α binding DNA in the absence of HIF-1 β .

I also wished to use the acquired data to explore whether HIF-1 β bound DNA at sites distinct from HIF- α in the tested cell lines since HIF-1 β is known to bind chromatin in association with alternate dimerisation partners such as AhR or SIM1/2 (Kewley et al., 2004). Although not in themselves hypoxia-inducible, it is possible that association of HIF-1 β with other factors may limit the HIF response by restricting HIF-1 β availability. I therefore undertook a quantitative assessment of the ability of HIF-1 β to bind DNA in the absence of HIF- α

2.2 Results

2.2.1 Assessment of concordance between ChIP-seq biological replicates

ChIP-seq data was generated in biological replicate in three separate cell lines (the hepatoma cell line, HepG2; the renal proximal tubule cell line, HKC8; and the VHL-defective ccRCC cell line, RCC4) for each HIF isoform (HIF-1 α , HIF-2 α and HIF-1 β). HepG2 and HKC8 cells were cultured in an atmosphere of 0.5% O₂ for 16 hours prior to ChIP-seq analysis, whilst experiments using the VHL-defective RCC4 cells were performed in normoxic conditions. ChIP-seq peaks were called by the peak caller, MACS (Zhang et al., 2008b). To confidently term a 'called peak' a 'transcription factor binding site', said regions must be called as peaks in both biological replicates. In order to identify a sufficient number of HIF binding sites to conduct meaningful downstream analysis, ChIP-seq replicates are required to be concordant with one another.

To quantify concordance between biological replicate experiments, all HIF-1 β binding sites in each of the three cell lines were first determined by identifying all MACS-called peaks that overlapped by one or more base pairs via the use of BEDTools intersect (v2.26.0). Peaks identified in replicate one that overlapped a peak in replicate two were maintained for further analysis. Within the HIF-1 β bound regions, HIF-1 α , HIF-2 α and HIF-1 β RPKM (read counts per kilobase per million mapped reads) was quantified for each replicate in the corresponding cell line. For each pair of ChIP-seq replicates in each cell line, scatter plots were generated and the Spearman's rank correlation coefficient was quantified (Figure 1A-I). All HIF ChIP-seq replicates in each of the assayed cell lines display highly concordant signals, with no R² value being below 0.75, therefore all ChIP-seq data displayed here was henceforth considered suitable for further analyses.

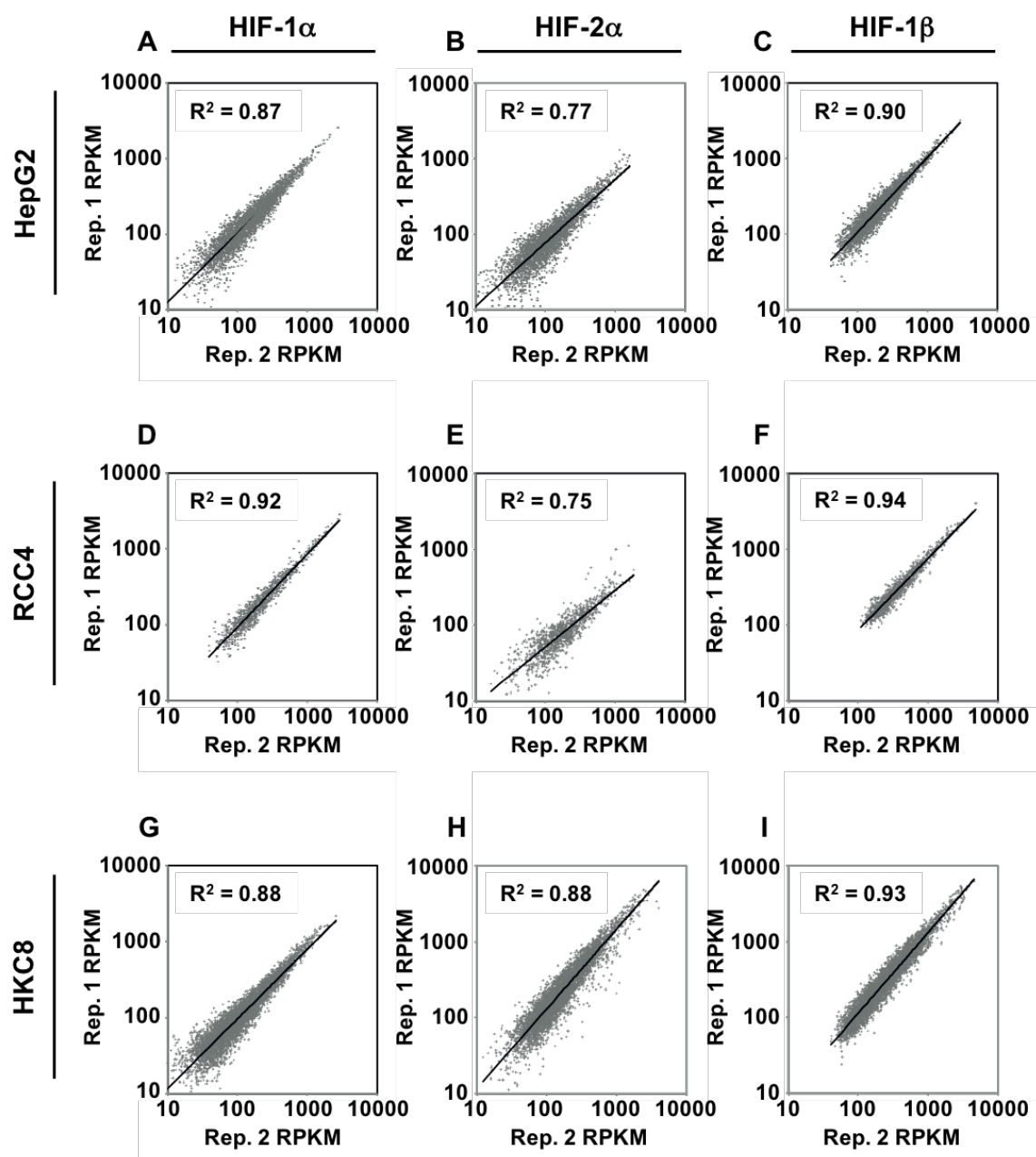


Figure 1. HIF ChIP-seq replicate concordance.

(A-C) Scatter plots of HIF-1 α , HIF-2 α and HIF-1 β RPKM signals for the HIF-1 β bound regions in HepG2 cells. R^2 = the Spearman's rank correlation coefficient.

(D-F) Scatter plots of HIF-1 α , HIF-2 α and HIF-1 β RPKM signals for the HIF-1 β bound regions in RCC4 cells. R^2 = the Spearman's rank correlation coefficient.

(G-I) Scatter plots of HIF-1 α , HIF-2 α and HIF-1 β RPKM signals for the HIF-1 β bound regions in HKC8 cells. R^2 = the Spearman's rank correlation coefficient.

2.2.2 Peak calling concordance, and statistical assessment of ChIP-seq signal concordance between HIF- α and HIF-1 β

ChIP-seq peak callers, such as MACS, are designed to call true positive transcription factor binding sites (Nakato and Shirahige, 2016) and therefore by necessity may falsely call negative binding at other regions. Therefore, while they may call high confidence binding sites, it is likely that some regions will not be called as peaks because they do not surpass the signal threshold needed to be called a peak. These peaks may nevertheless still represent a transcription factor binding site, albeit one of low affinity. In light of this, it is not possible to categorically state that a HIF- α peak not intersecting a HIF-1 β peak in the same cell line represents a HIF- α binding site existing in the absence of concurrent HIF-1 β binding (non-canonical HIF- α binding site).

To illustrate this issue, I have produced Venn diagrams to show the intersect between HIF- α and HIF-1 β peaks in HepG2 cells (Figure 2A). All sets of HIF peaks were high confidence peaks that had been called by MACS in both ChIP-seq biological replicates. 1,813 peaks are called in both HIF-1 α and HIF-1 β ChIP-seq experiments. However, 159 peaks are called in the HIF-1 α dataset alone, and 1,647 are deemed specific to HIF-1 β . These data suggest that HIF-1 α is capable of binding 159 sites in the absence of HIF-1 β , however it is possible that HIF-1 β is bound at these 159 regions but HIF-1 β reads at these sites merely did not surpass the MACS peak-calling threshold. To test this hypothesis, I plotted the mean (of two replicates) HIF-1 α signal (RPKM) against the mean HIF-1 β signal within both HIF-1 α peaks that overlapped a called HIF-1 β site and within those not-overlapping a called HIF-1 β peak, (Figure 2B-C). As expected,

there was a strong positive correlation (Spearman's rank correlation coefficient, $R^2 = 0.86$) between HIF-1 α and HIF-1 β RPKM values within peaks identified by the MACS peak caller as co-occupied by both HIF-1 α and HIF-1 β . However, contrary to the hypothesis that HIF-1 α peaks not intersecting HIF-1 β called peaks do not have HIF-1 β signal, HIF-1 α and HIF-1 β signals still exhibited a strong correlation even within HIF-1 α peaks that do not intersect a HIF-1 β peak ($R^2 = 0.58$).

I next examined the concordance of HIF-2 α and HIF-1 β called peaks in the same cell line. Markedly similar results to those seen for HIF-1 α data were observed. 149 peaks are called in HIF-2 α ChIP-seq samples that are not called in HIF-1 β (Figure 2D). 1,001 sites are common to both HIF-2 α and HIF-1 β . Again, similar to the concordant relationship between HIF-1 α and HIF-1 β , HIF-2 α and HIF-1 β exhibit a strong positive correlation in their RPKM signals within both the 1,001 sites common to HIF-2 α and HIF-1 β ($R^2 = 0.61$; Figure 2E), and in the 149 sites specifically called in the HIF-2 α ChIP-seq experiments alone ($R^2 = 0.49$; Figure 2F). These data suggest that peaks called in either endogenous HIF-1 α or HIF-2 α ChIP-seq experiments but not in HIF-1 β , do indeed have HIF-1 β signal that correlates in intensity with the HIF- α signal. However, on average, these regions have lower affinity and are therefore lower confidence binding sites, and are thus less likely to be called as a peak by the MACS peak caller.

Endogenous HIF-1 α and HIF-2 α in HepG2 cells appear to exhibit no obvious propensity to bind a significant amount of sites in a non-canonical fashion. However, to establish with confidence that this is a phenomenon conserved for both HIF- α subunits across other cell types, I analysed HIF-1 α ,

HIF-2 α and HIF-1 β ChIP-seq data in a further two cell lines in an identical manner to the previously described HepG2 analyses.

Analysis of both RCC4 and HKC8 ChIP-seq data yielded similar results to those observed in HepG2 cells. In RCC4 cells, a slightly greater proportion of HIF-1 α and HIF-2 α sites are called specifically in the HIF- α ChIP-seq experiments but not in that of HIF-1 β . Specifically, 383 sites are called HIF-1 α alone, whilst 519 are called in both HIF-1 α and HIF-1 β (Figure 3A). Furthermore, 605 sites are called in HIF-2 α alone, whilst 611 sites are called in both HIF-2 α and HIF-1 β (Figure 3D). Conversely, in HKC8 cells, only 33 sites are called HIF-1 α alone, whereas 1,083 sites are called as both HIF-1 α and HIF-1 β binding (Figure 4A). 617 sites are called in HIF-2 α alone, whereas 3,246 sites are called in both HIF-2 α and HIF-1 β datasets (Figure 4D).

As in HepG2 cells, HIF- α sites that do and do not overlap a called HIF-1 β peak in both RCC4 and HKC8 cells, exhibit a strong positive correlation in HIF- α and HIF-1 β RPKM. The Spearman's rank correlation coefficient determined for HIF- α sites that are and are not overlapping a called HIF-1 β peak are highly similar. The calculated R^2 value for the correlation between HIF-1 α and HIF-1 β signal within the totality of overlapping peaks ($R^2 = 0.77$) is essentially identical to that of the 383 HIF-1 α peaks not overlapping a HIF-1 β peak ($R^2 = 0.77$) (Figure 3B-C). Again, there was a remarkably similar degree of correlation between that of the 611 HIF-2 α sites in RCC4 cells that are overlapping a HIF-1 β peak ($R^2 = 0.39$), as compared to the 605 HIF-2 α sites in RCC4 cells that are not overlapping a HIF-1 β peak ($R^2 = 0.43$) (Figure 3E-F).

Comparable findings were observed in the concordance between HIF- α and HIF-1 β ChIP-seq signals in HKC8 cells. For the 1,083 HIF-1 α sites that are overlapping a HIF-1 β peak, the $R^2 = 0.76$, whilst for the 33 HIF-1 α sites in HKC8 cells that are not overlapping a HIF-1 β peak, the $R^2 = 0.93$ (Figure 4B-C). For the 3,246 HIF-2 α sites in HKC8 cells that are overlapping a HIF-1 β peak, the $R^2 = 0.79$, whilst for the 617 HIF-2 α sites in HKC8 cells that are not overlapping a HIF-1 β peak, the $R^2 = 0.60$ (Figure 4E-F).

Taking the RCC4 and HKC8 data together, the results corroborate those observed in HepG2 cells, in that even where called HIF- α peaks are not overlapping a called HIF-1 β peak, HIF-1 β is still likely to be co-occupying the HIF- α binding site. In many of these instances it is likely that the HIF-1 β reads are below the required signal threshold to be called a peak by MACS.

In contradistinction, the heatmap analysis shown later in this chapter (Figure 6) indicates that overexpressed exogenous HIF-1 α and HIF-2 α in 786O (another VHL-defective ccRCC cell line) cells appear to be capable of binding a significant set of sites in the absence of HIF-1 β . I would therefore anticipate that, unlike endogenous HIF- α in HepG2, RCC4 and HKC8 cells, exogenous HIF- α ChIP-seq experiments in 786O cells would lead to a significant number of called peaks that are not overlapping a called HIF-1 β peak. Moreover, I would expect that exogenous HIF- α peaks that do not overlap a HIF-1 β peak, would harbour a weaker correlation between HIF- α and HIF-1 β ChIP-seq signals.

The data described here has been derived from published ChIP-seq datasets (Salama et al., 2015). The 786O cells used in this published study had been stably transduced with the lentiviral overexpression constructs, pRRL empty

vector, pRRL HIF-1 α or pRRL HIF-2 α . All ChIP-seq experiments were performed in normoxia because HIF- α subunits are constitutively stabilised by the lack of functional VHL in 786O cells. Raw data derived from the single replicate ChIP-seq experiments were processed as previously described (Salama et al., 2015). Due to these experiments being performed in singlicate, only peaks called by each of two independent peak callers, T-PIC (Hower et al., 2011) and MACS (Zhang et al., 2008b), were considered.

A Venn diagram illustrates that 2,402 re-expressed full-length HIF-1 α (RE) sites in 786O pRRL HIF-1 α cells overlap HIF-1 β binding sites, whereas 2,341 sites are called in the HIF-1 α (RE) ChIP-seq experiment alone (Figure 5A). The 2,402 HIF-1 α (RE) peaks that are overlapping HIF-1 β peaks exhibit a positive correlation between HIF-1 α and HIF-1 β RPKM ($R^2 = 0.39$; Figure 5B). However, no correlation between HIF-1 α and HIF-1 β ChIP-seq signal was observed for the HIF-1 α (RE) sites that do not overlap a HIF-1 β peak ($R^2 = 0.04$; Figure 5C).

A Venn diagram illustrates that 2,055 overexpressed HIF-2 α (OE) sites in 786O pRRL HIF-2 α cells overlap HIF-1 β binding sites, whereas 3,201 sites are called in the HIF-2 α (OE) ChIP-seq experiment alone (Figure 5D). Analogous to HIF-1 α (RE), the 2,055 HIF-2 α (OE) peaks that are overlapping HIF-1 β peaks exhibit a positive correlation between HIF-2 α and HIF-1 β RPKM ($R^2 = 0.39$; Figure 5E). However, the HIF-2 α (OE) sites that do not overlap a HIF-1 β peak exhibit no correlation between HIF-2 α and HIF-1 β ChIP-seq signal ($R^2 = 0.05$; Figure 5F).

Analysis of endogenous HIF-2 α ChIP-seq data in 786O pRRL cells reveals intermediate results between those of exogenous HIF- α in 786O cells

and endogenous HIF- α in HepG2, RCC4 and HKC8 cells. Whilst the majority of the called HIF-2 α peaks in 786O pRRL cells overlap HIF-1 β , the correlation between HIF-2 α and HIF-1 β RPKM within HIF-2 α peaks that are not overlapping a HIF-1 β peak is somewhat weaker. 3,246 HIF-2 α sites in 786O pRRL cells overlap HIF-1 β binding sites, whereas 617 sites are called as HIF-2 α peaks but not HIF-1 β (Figure 5G). The 3,246 HIF-2 α sites that overlap HIF-1 β peaks exhibit a positive correlation in RPKM between HIF-2 α and HIF-1 β ($R^2 = 0.37$; Figure 5H). However, the 617 HIF-2 α sites that do not overlap HIF-1 β peaks exhibit a weak positive correlation ($R^2 = 0.18$; Figure 5I). Notably, this is not as weak a correlation as is observed among the exogenous overexpressed HIF-1 α and HIF-2 α ChIP-seq data (Figure 5C; 5F), however it is not nearly as strong as the correlation observed for the endogenous HIF- α subunits in HepG2, RCC4 and HKC8 cells.

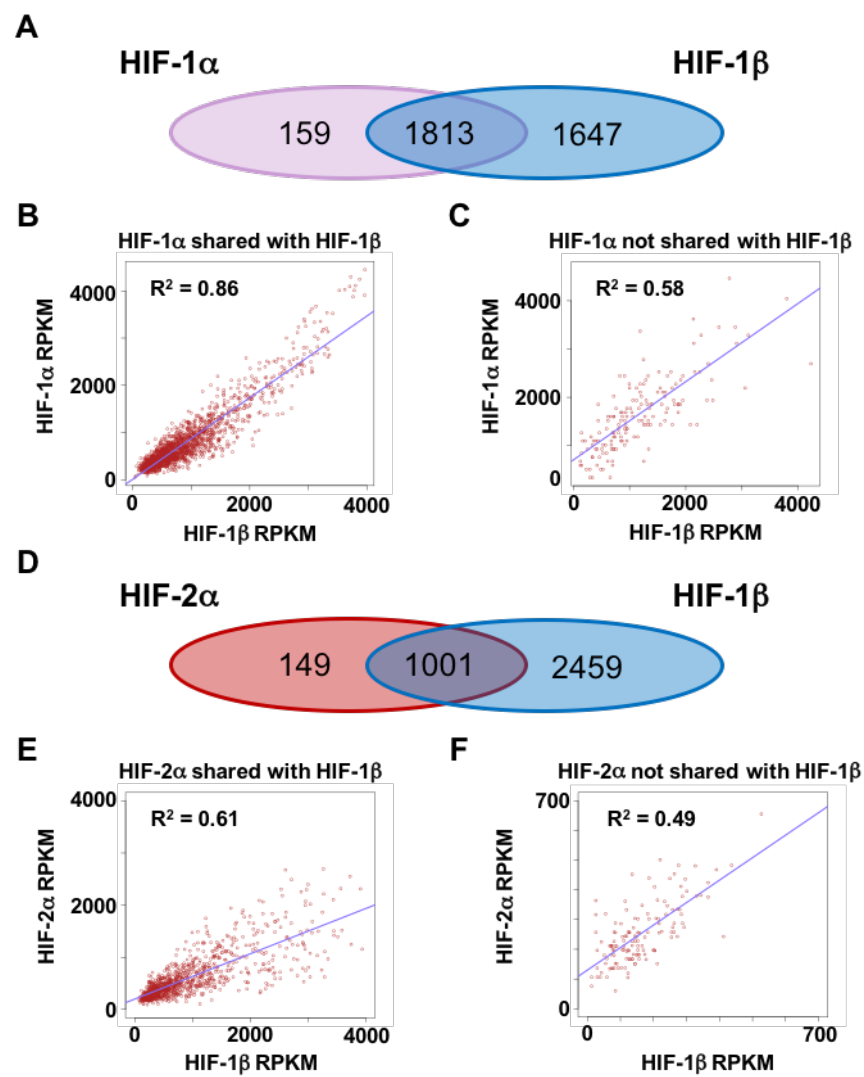


Figure 2. HIF-1 β and HIF- α concordance in HepG2 cells.

(A) Venn diagram illustrating the number of co-occupied and distinct HIF-1 α and HIF-1 β sites in HepG2 cells, as determined by MACS peak caller analysis of biological duplicate ChIP-seq assays.

(B) Scatter plot of HIF-1 α and HIF-1 β RPKM signals for the 1,813 HIF-1 α peaks, called in biological duplicate ChIP-seq samples that are overlapping a HIF-1 β peak. R^2 = the Spearman's rank correlation coefficient.

(C) Scatter plot of HIF-1 α and HIF-1 β RPKM signals for the 159 HIF-1 α peaks, called in biological duplicate ChIP-seq samples that are not overlapping a HIF-1 β peak.

(D) Venn diagram illustrating the number of co-occupied and distinct HIF-2 α and HIF-1 β sites in HepG2 cells, as determined via biological replicate ChIP-seq assays.

(E) Scatter plot of HIF-2 α and HIF-1 β RPKM signals for the 1,001 HIF-2 α peaks, called in biological duplicate ChIP-seq samples that are overlapping a HIF-1 β peak.

(F) Scatter plot of HIF-2 α and HIF-1 β RPKM signals for the 149 HIF-2 α peaks, called in biological duplicate ChIP-seq samples that are not overlapping a HIF-1 β peak.

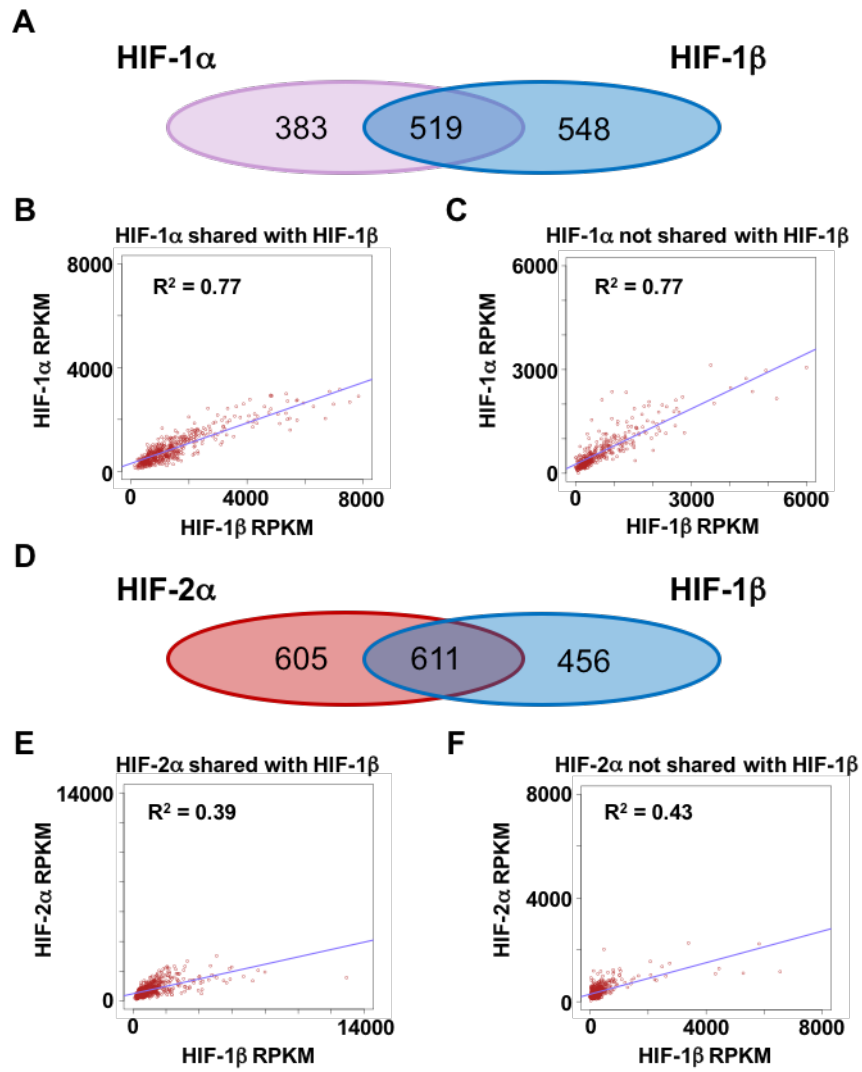


Figure 3. HIF-1 β and HIF- α concordance in RCC4 cells.

(A) Venn diagram illustrating the number of co-occupied and distinct HIF-1 α and HIF-1 β sites in RCC4 cells, as determined by MACS peak caller analysis of biological duplicate ChIP-seq assays.

(B) Scatter plot of HIF-1 α and HIF-1 β RPKM signals for the 519 HIF-1 α peaks, called in biological duplicate ChIP-seq samples that are overlapping a HIF-1 β peak. R^2 = the Spearman's rank correlation coefficient.

(C) Scatter plot of HIF-1 α and HIF-1 β RPKM signals for the 383 HIF-1 α peaks, called in biological duplicate ChIP-seq samples that are not overlapping a HIF-1 β peak.

(D) Venn diagram illustrating the number of co-occupied and distinct HIF-2 α and HIF-1 β sites in RCC4 cells, as determined via biological replicate ChIP-seq assays.

(E) Scatter plot of HIF-2 α and HIF-1 β RPKM signals for the 611 HIF-2 α peaks, called in biological duplicate ChIP-seq samples that are overlapping a HIF-1 β peak.

(F) Scatter plot of HIF-2 α and HIF-1 β RPKM signals for the 605 HIF-2 α peaks, called in biological duplicate ChIP-seq samples that are not overlapping a HIF-1 β peak.

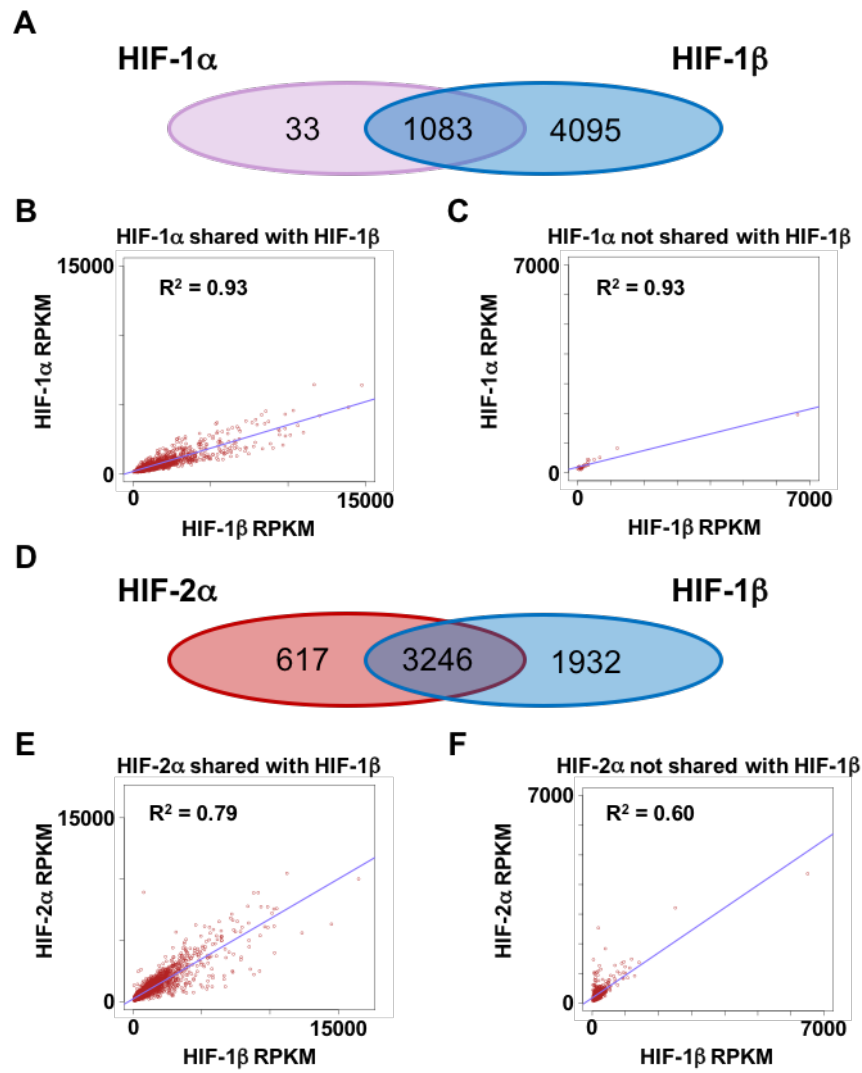


Figure 4. HIF-1 β and HIF- α concordance in HKC8 cells.

(A) Venn diagram illustrating the number of co-occupied and distinct HIF-1 α and HIF-1 β sites in HKC8 cells, as determined by MACS peak caller analysis of biological duplicate ChIP-seq assays.

(B) Scatter plot of HIF-1 α and HIF-1 β RPKM signals for the 1083 HIF-1 α peaks, called in biological duplicate ChIP-seq samples that are overlapping a HIF-1 β peak. R^2 = the Spearman's rank correlation coefficient.

(C) Scatter plot of HIF-1 α and HIF-1 β RPKM signals for the 33 HIF-1 α peaks, called in biological duplicate ChIP-seq samples that are not overlapping a HIF-1 β peak.

(D) Venn diagram illustrating the number of co-occupied and distinct HIF-2 α and HIF-1 β sites in HKC8 cells, as determined via biological replicate ChIP-seq assays.

(E) Scatter plot of HIF-2 α and HIF-1 β RPKM signals for the 3246 HIF-2 α peaks, called in biological duplicate ChIP-seq samples that are overlapping a HIF-1 β peak.

(F) Scatter plot of HIF-2 α and HIF-1 β RPKM signals for the 617 HIF-2 α peaks, called in biological duplicate ChIP-seq samples that are not overlapping a HIF-1 β peak.

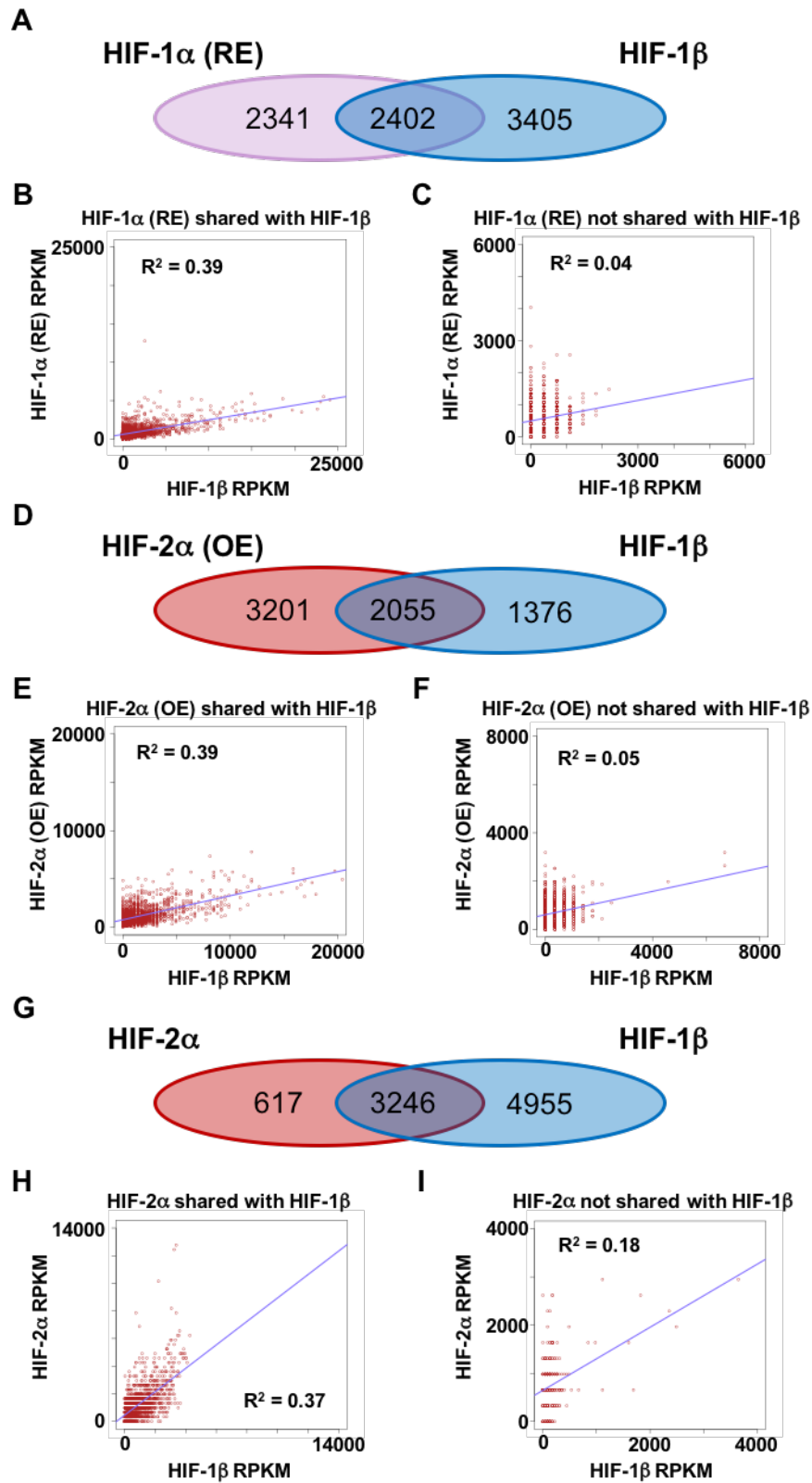


Figure 5. HIF-1 β and HIF- α concordance in 7860 cells.

(A) Venn diagram illustrating the number of co-occupied and distinct HIF-1 α (RE) and HIF-1 β sites in 7860 pRRL HIF-1 α cells, as determined by MACS peak caller analysis of biological duplicate ChIP-seq assays.

(B) Scatter plot of HIF-1 α (RE) and HIF-1 β RPKM signals for the 2402 HIF-1 α (RE) peaks, called in biological duplicate ChIP-seq samples that are overlapping a HIF-1 β peak. R^2 = the Spearman's rank correlation coefficient.

(C) Scatter plot of HIF-1 α (RE) and HIF-1 β RPKM signals for the 2341 HIF-1 α (RE) peaks, called in biological duplicate ChIP-seq samples that are not overlapping a HIF-1 β peak.

(D) Venn diagram illustrating the number of co-occupied and distinct HIF-2 α (OE) and HIF-1 β sites in 7860 pRRL HIF-2 α cells, as determined by MACS peak caller analysis of biological duplicate ChIP-seq assays.

(E) Scatter plot of HIF-2 α (OE) and HIF-1 β RPKM signals for the 2055 HIF-2 α (OE) peaks, called in biological duplicate ChIP-seq samples that are overlapping a HIF-1 β peak.

(F) Scatter plot of HIF-2 α (OE) and HIF-1 β RPKM signals for the 3201 HIF-2 α (OE) peaks, called in biological duplicate ChIP-seq samples that are not overlapping a HIF-1 β peak.

(G) Venn diagram illustrating the number of co-occupied and distinct endogenous HIF-2 α and HIF-1 β sites in 7860 pRRL cells, as determined by MACS peak caller analysis of biological duplicate ChIP-seq assays.

(H) Scatter plot of HIF-2 α and HIF-1 β RPKM signals for the 3246 HIF-2 α peaks, called in biological duplicate ChIP-seq samples that are overlapping a HIF-1 β peak.

(I) Scatter plot of HIF-2 α and HIF-1 β RPKM signals for the 617 HIF-2 α peaks, called in biological duplicate ChIP-seq samples that are not overlapping a HIF-1 β peak.

2.2.3 Re-assessment of published endogenous and exogenous HIF-1 α and HIF-2 α ChIP-seq data in 786O cells, and association of these data with matched analyses of endogenous HIF-1 α and HIF-2 α ChIP-seq in other cell lines

In a study published by the Ratcliffe group in 2015 (Salama et al., 2015), heatmap analyses were performed using ChIP-seq data from 786O cells in order to compare HIF- α and HIF-1 β binding. An absence of HIF-1 β was observed at a significant proportion of HIF-2 α binding sites when HIF-2 α was overexpressed (OE) approximately 8-fold in 786O cells. The observed non-canonical binding of HIF-2 α was reported to be in contrast to exogenously re-expressed (RE) HIF-1 α that did not appear to bind sites without HIF-1 β . However, interpretations were made based on heatmap analyses in which HIF-1 α (RE) and HIF-2 α (OE) peak summits (peak summits are 1 bp coordinates within ChIP-seq peaks whereby the ChIP-seq signal for the factor being subjected to ChIP is at its highest) were sorted by the ChIP-seq signal intensity of the HIF- α subunit, rather than HIF-1 β . This made identification of HIF- α sites lacking detectable HIF-1 β signal more difficult than it might have been had the HIF- α peak summits been sorted by HIF-1 β ChIP-seq signal. I therefore re-analysed the data from this study, but now ranked peaks according to HIF-1 β signal.

As expected there is strong HIF-1 α ChIP-seq signal at all sites identified as binding HIF-1 α by T-PIC and MACS in this dataset (Figure 6A). However, HIF-1 β ChIP-seq signal is less consistently observed at HIF-1 α sites. In particular approximately half of the HIF-1 α sites show little to no HIF-1 β signal above local

background levels. This is consistent with widespread binding of exogenous HIF-1 α in the absence of HIF-1 β .

A comparable analysis of the capacity of exogenous HIF-2 α to bind chromatin in the absence of HIF-1 β , was then performed using HIF-2 α (OE) and HIF-1 β ChIP-seq datasets from 786O pRRL HIF-2 α cells (Figure 6B). Similarly, about half of the HIF-2 α sites again had little to no HIF-1 β signal when compared to local background. However, when the analysis was repeated for the endogenous HIF-2 α and HIF-1 β ChIP-seq datasets in 786O pRRL cells (Figure 6C) strong HIF-1 β signal was observed across almost all endogenous HIF-2 α sites. This suggests that the observed HIF- α binding in the absence of HIF-1 β in the cell lines overexpressing HIF- α subunits may be the result of non-physiological expression of HIF- α .

I therefore wished to determine whether HIF- α binding in the absence of HIF-1 β might be determined under physiological conditions and in particular whether it might be observed in other cell types. I therefore analysed endogenous HIF ChIP-seq experiments to assay pan-genomic binding of endogenous HIF- α subunits, alongside HIF-1 β , in HepG2, HKC8, and RCC4 cells. All endogenous ChIP-seq experiments were performed in biological duplicate.

ChIP-seq peaks were called using the MACS peak caller, and only those identified in both biological replicates were considered. Specifically, peaks were first called in each replicate independently. Peaks from replicate 1 that overlapped a peak in replicate 2 by one or more base pairs were then identified using of BEDTools intersect (v2.26.0) and taken forward for further analysis. HIF-

1 α and HIF-2 α peak summits were sorted, as before, according to HIF-1 β ChIP-seq signal intensity (mean RPKM across the two biological replicates) within the given summit regions. Heatmaps were again generated to display HIF-1 β and HIF- α ChIP-seq signal intensities (CPM, read counts per million mapped reads) on a colour scale using ngs.plot (Figure 7A-C).

In each cell line strong HIF-1 β ChIP-seq signal was observed across almost all HIF-1 α and HIF-2 α binding sites. Taken together with the re-analysis of endogenous HIF-2 α and HIF-1 β binding in 786O cells, this provides little evidence to support the possibility of widespread non-canonical binding of HIF- α subunits in the absence of HIF-1 β in any of the 4 cell lines examined. This is in contradistinction to 786O cell lines overexpressing exogenous HIF-1 α or HIF-2 α .

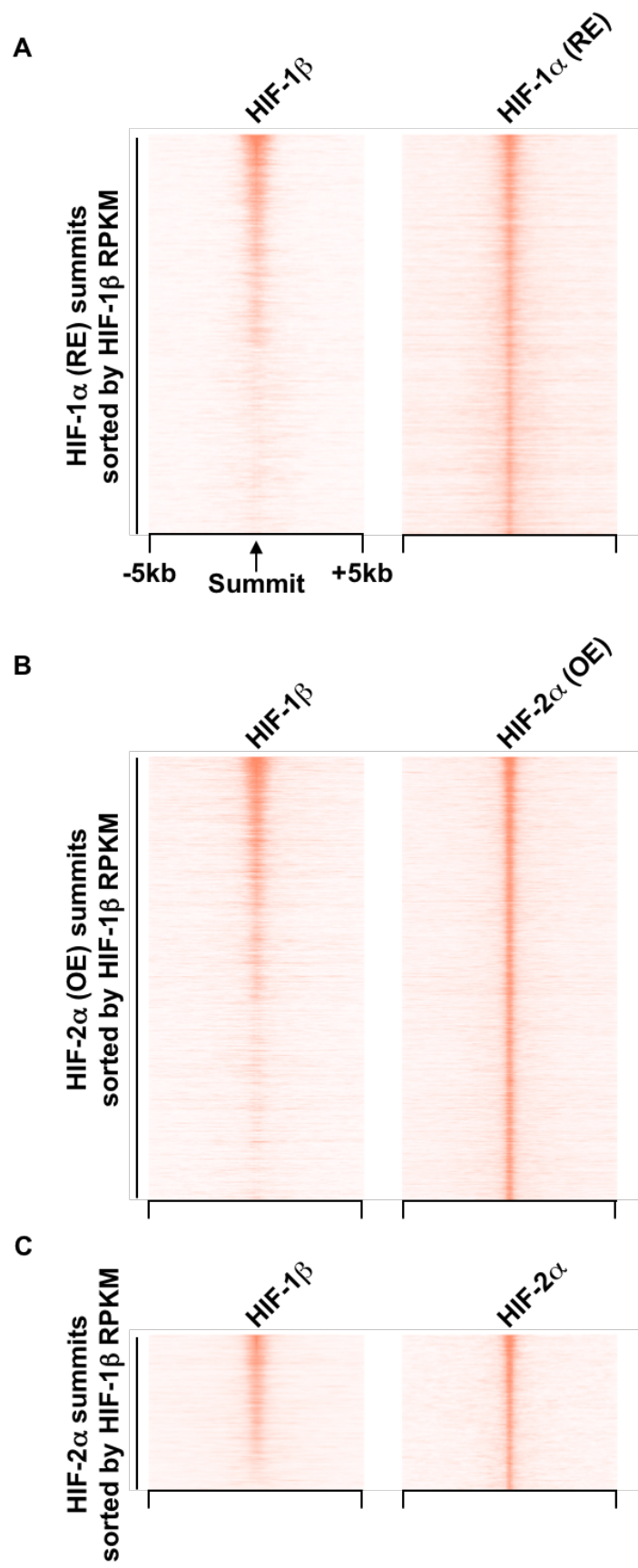


Figure 6. HIF- α and HIF-1 β binding site co-occupancy in 786O cells.

Heatmaps showing local HIF-1 β , HIF-1 α and HIF-2 α ChIP-seq signal intensity within ± 5 kb regions around peak summits (read counts per million mapped reads (CPM) expressed as colour intensity). Peaks have been sorted on the vertical axis according to HIF-1 β RPKM (average read counts per kilobase per million mapped reads across two biological replicates). The x-axis represents the distance from each peak summit.

(A) Signal intensity for HIF-1 β and re-expressed (RE) HIF-1 α around peak summits for HIF-1 α (RE) sorted by HIF-1 β signal in 786O cells lentivirally transduced with pRRL HIF-1 α .

(B) Signal intensity for HIF-1 β and over-expressed (OE) HIF-2 α around peak summits for HIF-2 α (OE) sorted by HIF-1 β signal in 786O cells lentivirally transduced with pRRL HIF-2 α .

(C) Signal intensity for HIF-1 β and endogenous HIF-2 α around peak summits for HIF-2 α sorted by HIF-1 β signal in 786O cells lentivirally transduced with pRRL empty vector control.

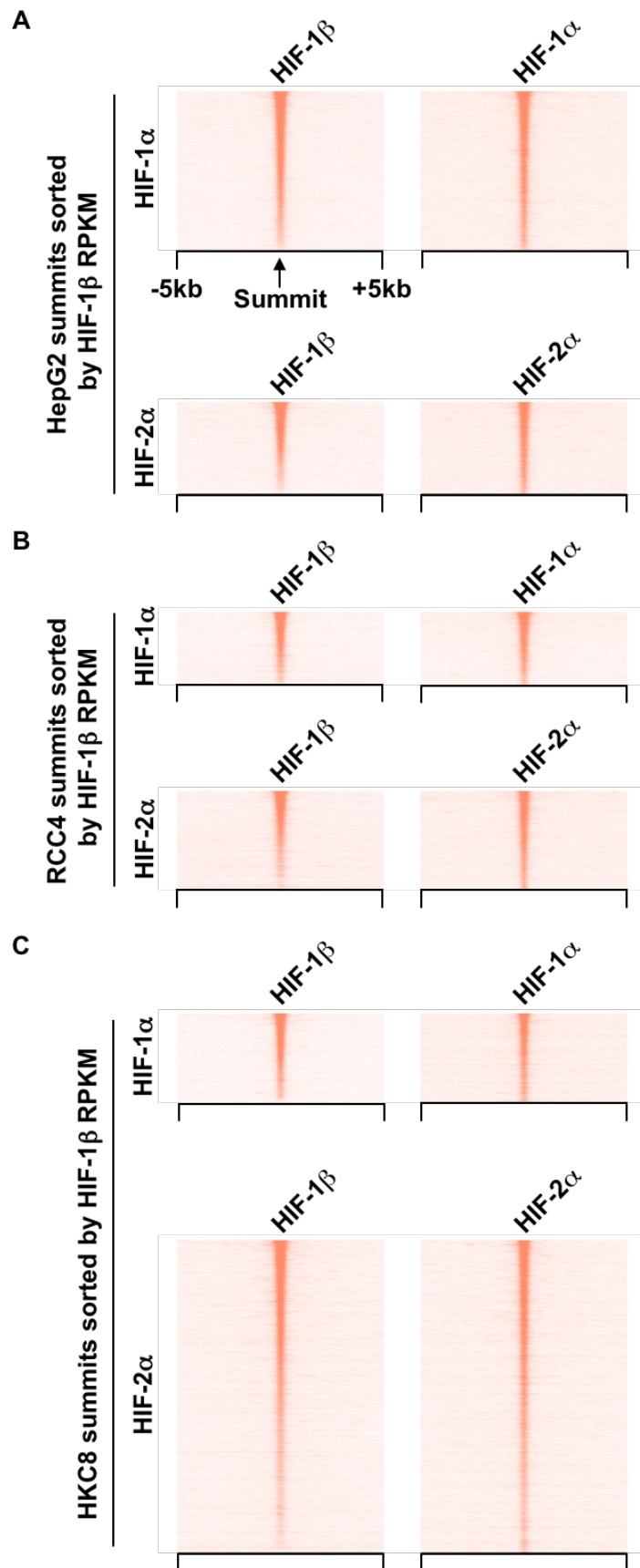


Figure 7. Endogenous HIF- α and HIF-1 β binding site co-occupancy in HepG2, RCC4 and HKC8 cells.

(A) Heatmaps showing HIF-1 β , HIF-1 α and HIF-2 α ChIP-seq signal (read counts per million mapped reads (CPM) expressed as colour intensity) for regions $\pm$ 5kb around HIF-1 α and HIF-2 α peak summits (x-axis) that have been sorted by HIF-1 β RPKM (vertical axis). ChIP-seq experiments were performed in biological duplicate on HepG2 cells incubated for 16 hours in 0.5% O₂. HIF- α peaks were only considered if called in both independent biological replicates. Merged .bam files were obtained by combining two biological replicates. These merged .bam files were used to map ChIP-seq signal intensity (CPM) across the defined regions depicted by the heatmaps.

(B) Heatmaps were generated as above for VHL-defective RCC4 cells incubated in normoxic conditions.

(C) Experiments were performed as above in HKC8 cells incubated for 16 hours in 0.5% O₂.

2.2.4 Further corroboration of the concordant relationship between endogenous HIF- α and HIF-1 β , with the exception of a small number of HIF-2 α binding sites

Whilst the analyses described above indicate the absence of widespread binding of endogenous HIF- α to chromatin without HIF-1 β , it is possible that a small number sites with a disproportionate amount of HIF- α to HIF-1 β signal exist.

To confirm the overall trends observed in the above analysis and to examine for specific exceptions to a general rule, I have undertaken an alternative quantitative analysis of HIF- α and HIF-1 β binding signals. Specifically, I have taken the high-confidence HIF-1 α and HIF-2 α peaks (peaks identified in both biological replicate ChIP experiments) in each cell line, and sorted these sites by the mean HIF-1 β RPKM for the two biological replicate ChIP-seq samples. I then generated line plots illustrating the mean HIF- α and HIF-1 β RPKM within these sorted peaks. Furthermore, I established a mean pan-genomic 'background' signal of HIF-1 β , with which to compare the HIF-1 β signal at HIF- α sites. Background signal was also plotted on the aforementioned line plots. This background signal is calculated using a multi-step process:

1. Using BEDTools intersect, I firstly established the genomic regions of each cell line wherein chromatin was accessible and also active. Open chromatin regions (as determined via DNase-seq or FAIRE-seq, depending on the cell line) were intersected with H3K27ac ChIP-seq peaks, whereby open chromatin regions were only further considered if they overlapped a H3K27ac peak by 1 or more bp (DNase-seq, FAIRE-seq and H3K27ac ChIP-seq data for all of the cell lines that I have

assayed are publically available. Accession numbers for each publically available dataset used are listed in the Appendix 8.1). This enabled identification of all genomic regions that are not only accessible for transcription factors to bind, but also harbour marks of active chromatin (Creyghton et al., 2010).

2. Using BEDTools subtract, I then subtracted all active accessible sites that overlap a HIF-1 β peak, and subsequently determined the mean HIF-1 β RPKM (using BEDTools multicov) within these remaining active and accessible regions of chromatin.
3. I then considered this value to be the mean 'background' HIF-1 β signal. This allows identification of regions in which there is significant HIF- α RPKM, but the concurrent HIF-1 β signal is equal to or below the level of HIF-1 β background signal.

The line plots indicate that endogenous HIF- α and HIF-1 β ChIP-seq signal in HepG2, RCC4 and HKC8 cells appear to be highly concordant with one another (Figure 8A-F). Within the vast majority of HIF- α binding sites, there is significant HIF-1 β signal that exceeds the level of the mean background HIF-1 β signal. This provides further corroboration of the previously described heatmap and scatter-plot analyses, thus further demonstrating the prevalence of pan-genomic canonical HIF- α binding sites. Likewise, the significant number of exogenous HIF- α binding sites in 786O cells with little to no HIF-1 β signal is consistent with the the previously discussed heatmap and scatter plot data (Figure 8G-I).

I then identified all HIF- α binding sites in HepG2, RCC4 and HKC8 cells in which HIF-1 β signal is less than the mean background signal of HIF-1 β . This led to identification of only 11 possible non-canonical HIF-2 α sites (2 in HepG2; 6 in RCC4; 3 in HKC8) and 1 possible non-canonical HIF-1 α site (in RCC4). Chromosome coordinates for all 12 sites are listed in the Appendix 8.3. and two representative loci are shown in Figures 9A and 9B.

To establish whether these 12 sites are likely to be functional, I interrogated whether the nearest gene to the non-canonical HIF- α binding site was significantly regulated in hypoxia using RNA-seq and microarray datasets. Unpublished microarray analysis of RCC4 cells reconstituted transfected with siRNA targeting HIF-1 α and HIF-2 α was used to identify hypoxic gene regulation in the RCC4 setting (generated by former Ratcliffe Group member, Raju Raval). RNA-seq analysis of HepG2 (my own data) and HKC8 cells (generated by former Ratcliffe Group member, Elizabeth Murray) cultured in normoxia or in 0.5% O₂ for 16 hours was used to identify hypoxic gene regulation in these cells. Next generation sequencing libraries were prepared from RNA by polyA⁺ selection, and RNA-seq was subsequently performed. RNA was harvested in biological triplicate for each condition. DESeq2 analysis was performed to determine the normoxic vs. hypoxic fold change across all detected transcripts, in addition to determining an adjusted *p* value as a measure of the statistical significance of transcript fold change. Of the 11 genes neighbouring each of the 12 non-canonical endogenous HIF- α binding sites, only *SLC25A17* was significantly regulated in hypoxia (0.81-fold regulation and *** *p* < 0.001 in HepG2 cells; Figure 9C), thus suggesting an overall lack of functionality of these sites, and suggesting that these “non-canonical sites” potentially represent noise.

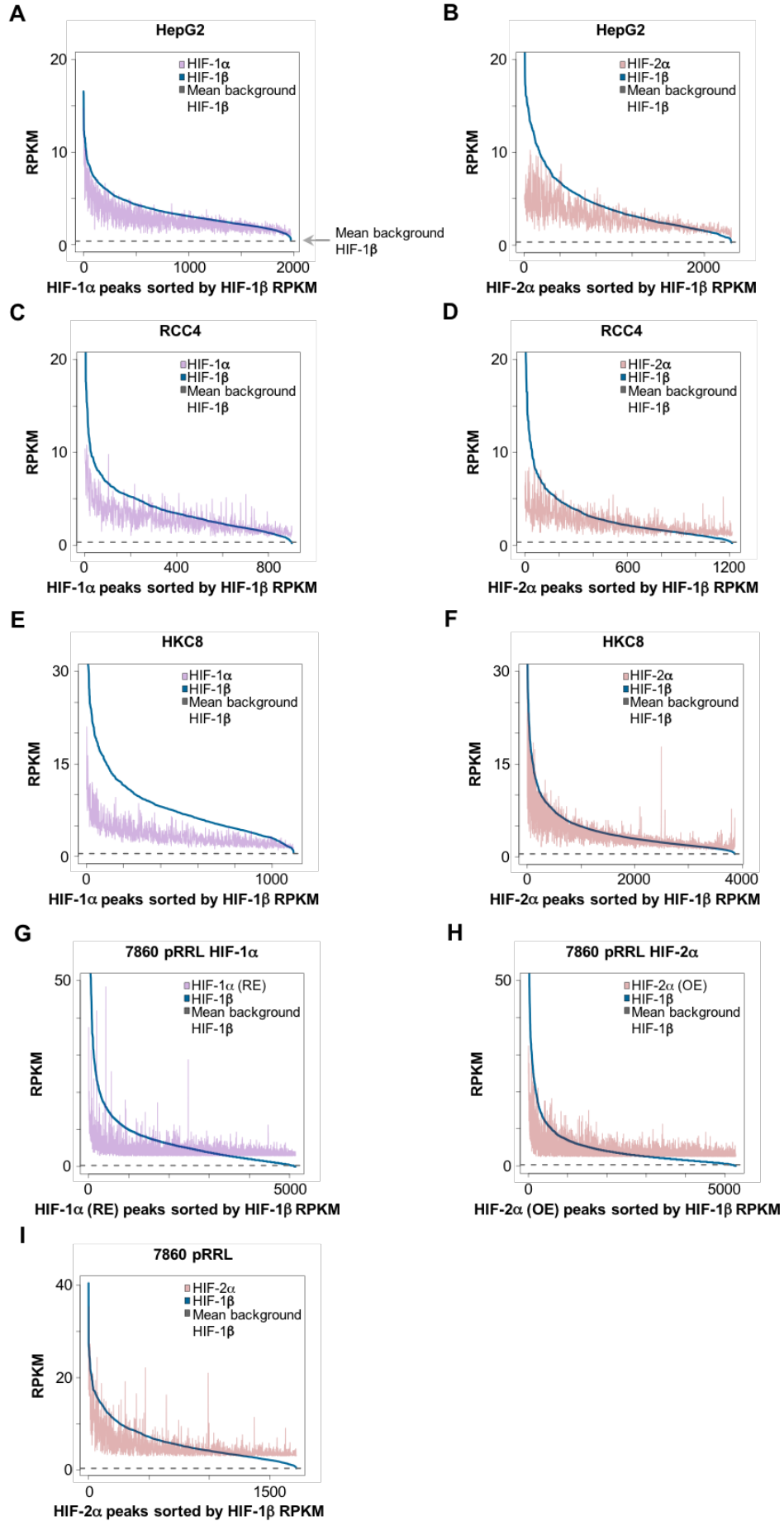


Figure 8. Identification of non-canonical HIF- α binding sites.

(A) Line plot showing MACS called HIF-1 α peaks in HepG2 cells that have been sorted by HIF-1 β signal at each peak. The RPKM of HIF-1 β within said peaks is represented by the blue line whereas HIF-1 α RPKM signal is represented by the purple line. The mean background signal, equates to the mean HIF-1 β RPKM across a random set of accessible regions that do not intersect HIF peaks.

(B) Line plot, as above, for MACS called HIF-2 α peaks sorted by HIF-1 β .

(C) Line plot, as above, for MACS called HIF-1 α peaks in RCC4 cells sorted by HIF-1 β .

(D) Line plot as above, for MACS called HIF-2 α peaks in RCC4 cells sorted by HIF-1 β .

(E) Line plot as above, for MACS called HIF-1 α peaks in HKC8 cells sorted by HIF-1 β .

(F) Line plot as above, for MACS called HIF-2 α peaks in HKC8 cells sorted by HIF-1 β .

(G) Line plot as above, for MACS called HIF-1 α (RE) peaks in 786O pRRL HIF-1 α cells sorted by HIF-1 β .

(H) Line plot as above, for MACS called HIF-2 α (OE) peaks in 786O pRRL HIF-2 α cells sorted by HIF-1 β .

(I) Line plot as above, for MACS called HIF-2 α peaks in 786O pRRL cells sorted by HIF-1 β .

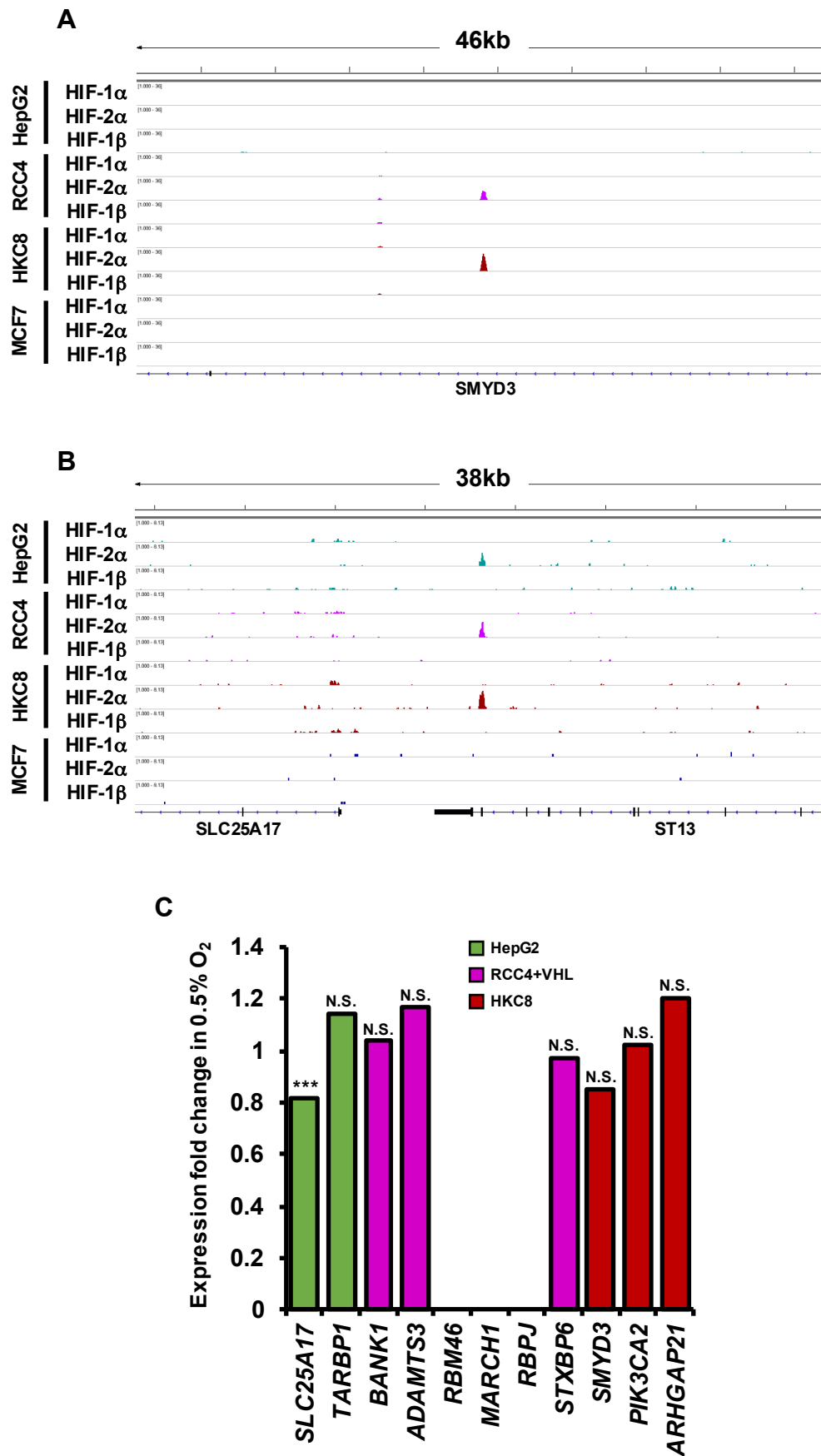


Figure 9. Non-canonical HIF- α binding sites.

(A) A 46 kb intronic genomic region in the *SMYD3* gene depicting representative HIF-1 α , HIF-2 α and HIF-1 β ChIP-seq signals in HepG2 (green), RCC4 (pink), HKC8 (red) and MCF7 (blue) cells.

(B) A 38 kb genomic region adjacent to the *SLC25A17* gene depicting representative HIF-1 α , HIF-2 α and HIF-1 β ChIP-seq signals in HepG2 (green), RCC4 (pink), HKC8 (red) and MCF7 (blue) cells.

(C) Bar chart illustrating the fold change in expression of 11 genes that are each adjacent to a non-canonical HIF- α binding site in either HepG2 (green), RCC4 (pink) or HKC8 (red) cells. Pan-genomic expression data was determined in RCC4 cells transfected with scrambled siRNA or a combination of siRNAs against HIF-1 α and HIF-2 α by performing microarray analysis. RNA-seq was used to determine pan-genomic gene expression in HepG2 and HKC8 cells. RNA was harvested in biological triplicate in cells cultured in normoxic conditions or for 16 hours in 0.5% O₂. Probes for *RBM46*, *MARCH1* and *RBPJ* were not present in the RCC4 microarray. *** $p < 0.001$; N.S. (not statistically significant) compared to normalised expression in normoxic conditions.

2.2.5 Exploring HIF-1 β binding sites for absence of HIF- α

Previously in this chapter, I have demonstrated that there is little evidence of widespread HIF- α binding in the absence of HIF-1 β . However, unlike HIF- α , HIF-1 β is known to have alternate dimerisation partners, including AhR (aryl hydrocarbon receptor) and SIM1 and 2 (single minded proteins) (Kewley et al., 2004). AhR is ligand-activated in response to dioxins and other related environmental pollutants. Upon ligand binding, AhR translocates to the nucleus and dimerises with HIF-1 β , via their respective PAS domains, and subsequently binds DNA to induce genes involved in metabolising xenobiotics (Lees and Whitelaw, 1999).

SIM1 and SIM2 have both previously been shown to constitutively interact with HIF-1 β , in a ligand-independent manner (Probst et al., 1997). Studies have demonstrated that expression patterns of SIM1 and SIM2 genes are restricted to a relatively small subset of mammalian tissues, including neuronal cells, the lung, and skeletal muscle, therefore the capacity of SIM1 and SIM2 to interact with HIF-1 β may be limited (Ema et al., 1996).

I therefore examined my ChIP-seq data in the three separate cell lines for the presence of HIF-1 β binding sites that lacked HIF- α signal. Reciprocal analyses to those described in Figure 8A-F of this chapter were performed. I first identified HIF-1 β binding sites in HepG2, RCC4 and HKC8 cells, using the MACS peak caller and BEDTools intersect (v2.26.0) (Dale et al., 2011) to determine peaks that were present in both biological replicates as before. Given that each HIF-1 β site may be bound by either HIF-1 α or HIF-2 α or by both isoforms, I wished to calculate the total HIF- α ChIP-seq signal at each HIF-1 β binding site. To do this, I first averaged the ChIP-seq signal (RPKM) for each HIF- α isoform

across the two replicates and then summed the signal for the two isoforms together at each HIF-1 β binding site. HIF-1 β binding sites were then ranked according to combined HIF- α signal and line plots illustrating mean HIF-1 β ChIP-seq signal (red) and HIF- α signal (blue) within the ranked HIF-1 β binding sites were generated as before (Figure 10A-C). The mean HIF- α 'background' signal was simultaneously calculated and plotted using the same methodology described previously in section 2.2.4.

Across HIF-1 β defined binding sites, the total HIF- α and HIF-1 β ChIP-seq signal in HepG2, RCC4 and HKC8 cells appear to be highly concordant with one another (Figure 10A-C). Furthermore, in all three cell lines, the HIF- α ChIP-seq signal exceeds background signal levels at the vast majority of HIF-1 β binding sites. This indicates that almost all HIF-1 β binding sites are co-occupied with HIF- α and again provides evidence against widespread binding of HIF-1 β in the absence of HIF- α in hypoxic HepG2, RCC4 and HKC8 cells.

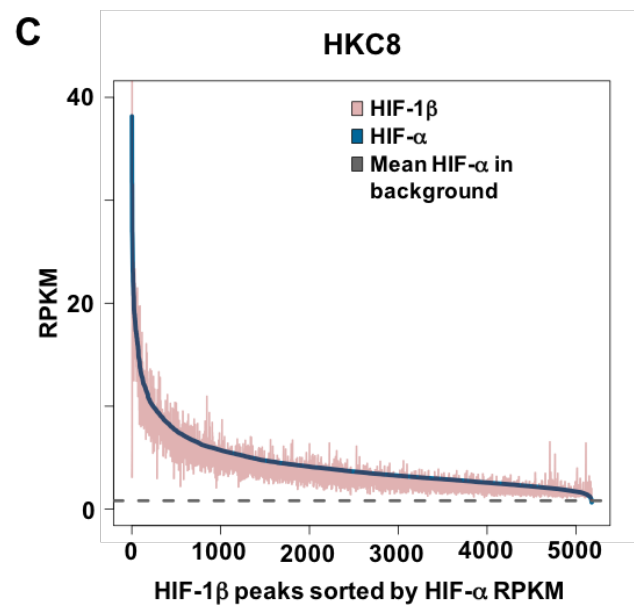
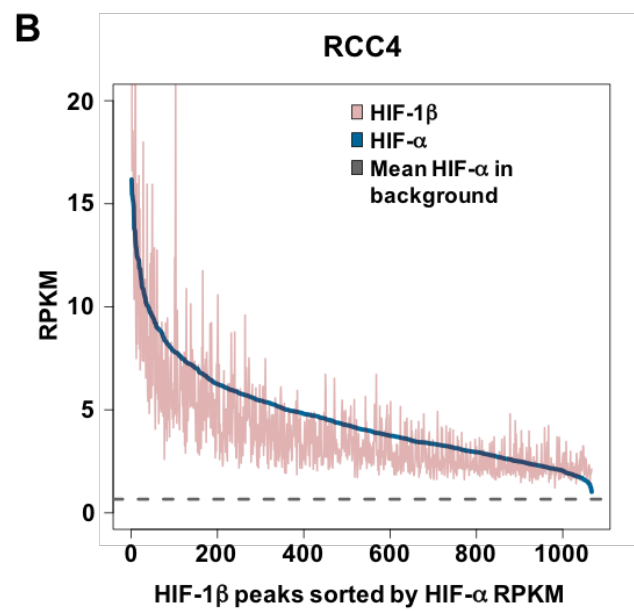
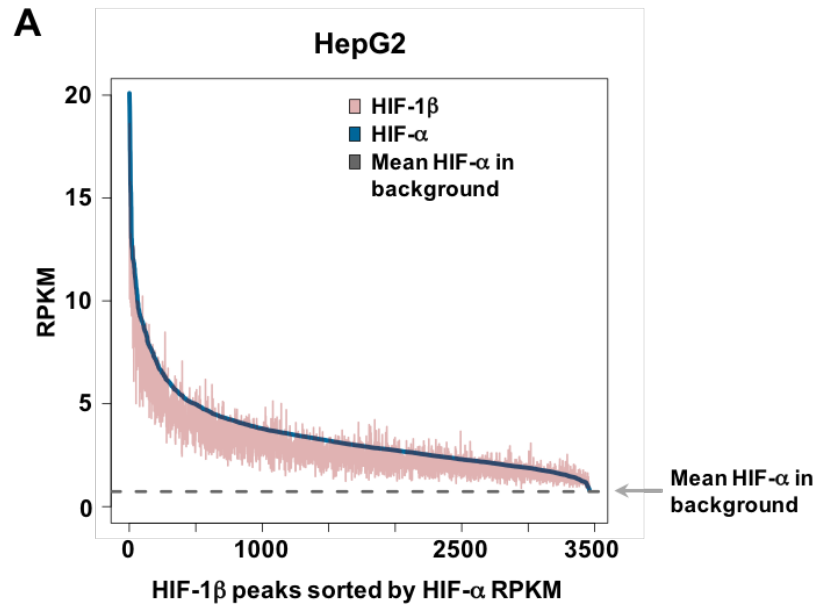


Figure 10. Exploring HIF-1 β sites for the absence of HIF- α signal.

(A) Line plot representing the totality of HIF-1 β binding sites in HepG2 cells that have been sorted by the summed RPKM of HIF-1 α and HIF-2 α within these peaks. The RPKM of HIF- α is represented by the blue line whereas HIF-1 β signal is represented by the red line. The mean background signal, equates to the mean HIF- α RPKM across a random set of accessible regions that do not intersect HIF peaks.

(B) Line plot as in (A), however HIF-1 β sites in RCC4 cells are sorted by HIF- α .

(C) Line plot as in (A), however HIF-1 β sites in HKC8 cells are sorted by HIF- α .

2.3 Discussion

This chapter describes the use of duplicate ChIP-seq data across a panel of cell lines in order to assess the levels of concordance between HIF- α and HIF-1 β binding signals. This has enabled a greater understanding of the extent of canonical versus non-canonical HIF- α binding than has previously been described in the literature (Gustafsson et al., 2005; Xiang et al., 2015; Salama et al., 2015).

Analysis of these ChIP-seq data sets has emphasised the difficulties involved in comparing ChIP-seq analyses, particularly when relying on peak callers to define sites as either binding or not binding in a binary fashion. Such methodologies clearly underestimate the degree of overlap between two datasets and although prevalent in the literature (Biddie et al., 2011; Zaret and Carroll, 2011; Ji et al., 2012; Grontved et al., 2013; Zanconato et al., 2015) should not be used without the addition of more quantitative methodologies.

2.3.1 HIF- α and HIF-1 β are largely concordant genome-wide

Re-analysis of 786O cell ChIP-seq data published by the Ratcliffe Group (Salama et al., 2015), in which HIF-1 α and HIF-2 α had been either re-introduced at high level or overexpressed, has shed new light upon the non-canonical activity of HIF-1 α and HIF-2 α in these cells. Although principal component analysis (PCA) in the published study shows that exogenous HIF-1 α signal aligns better with HIF-1 β than does HIF-2 α signal, I have observed following my re-analyses of this data that significant subsets of both exogenous HIF-1 α and HIF-2 α binding sites are lacking HIF-1 β .

Overexpression of a nuclear factor could feasibly lead to multiple off-target interactions with chromatin that are captured via the formaldehyde crosslinking step of the ChIP protocol (Solomon and Varshavsky, 1985). Salama et al., therefore provided evidence that argues against the possibility that the non-canonical binding sites of exogenous HIF- α are due to non-specific interaction. Firstly, the authors demonstrated that HIF-1 α (RE) and HIF-2 α (OE) binding sites were indeed functional and associated with induction of the nearest gene to the HIF- α binding site in RNA-seq analyses following overexpression of each HIF- α subunit. Secondly, the HIF-1 α (RE) and HIF-2 α (OE) binding sites were not randomly distributed across the genome, as would be expected if the binding sites were due to non-specific off-target chromatin associations. Specifically, the distribution of HIF-1 α (RE) and HIF-2 α (OE) sites with respect to the nearest transcriptional start site (TSS) was non-random and similar to that observed for intact endogenous HIF-1 and HIF-2 heterodimers in MCF7 cells (Schodel et al., 2011). In particular, HIF-1 α (RE) bound sites proximal to the TSS of the nearest gene, whereas HIF-2 α (OE) bound sites were more distal to the TSS of the nearest gene.

Salama et al., also examined for a trend in motif enrichment at HIF-1 α (RE) and HIF-2 α (OE) binding sites when sorted by HIF-1 β ChIP-seq signal. At HIF-1 α (RE) sites, the HRE was consistently enriched across all of the sorted HIF-1 α (RE) sites. However, at HIF-2 α (OE) sites, the HRE decreased in prevalence as HIF-1 β signal decreased, whilst enrichment of the AP-1 (TRE) motif progressively increased. The authors suggested that an interaction with AP-1 may be important for the non-canonical activity of HIF-2 α , for which there is

precedent in prior studies that have linked AP-1 to the functional activity of HIF (Laderoute, 2005; Villar et al., 2012). Conversely, since non-canonical HIF-1 α binding sites remain enriched for the HRE motif it might suggest that overexpressed HIF-1 α binds chromatin either as a monomer or in combination with an E-box binding dimerisation partner such as MYC or MAX. A third possibility that HIF-1 β might bind as a homodimer is less likely, since if this was the case, the DNA-binding motif would be unlikely to be a non-palindromic sequence such as the HRE.

Endogenous levels of HIF-1 β are generally thought to be in excess of HIF-1 α levels such that HIF-1 α is the limiting factor. Thus it is possible that at endogenous levels, HIF-1 α (whether free or bound to chromatin) only exists as a heterodimer with HIF-1 β . If this is the case, then HIF-1 α would always bind chromatin in complex with HIF-1 β since it would only be present in this form. However, if overexpressed at levels in excess of HIF-1 β , HIF-1 α might exist and therefore be able to bind chromatin without HIF-1 β . In this situation it is possible that overexpression of exogenous HIF-1 α might also compete with endogenous HIF-2 α for HIF-1 β and thereby free up HIF-2 α to bind without HIF-1 α . Thus it might be possible to see evidence of non-canonical HIF-2 α binding in HIF-1 α overexpressing cells.

2.3.2 Endogenous HIF ChIP-seq in HepG2, RCC4, HKC8 and 786O, and concordance of HIF-1 α and HIF-1 β

Although non-canonical binding of endogenous HIF-1 α subunits without HIF-1 β has not been observed in 786O or MCF7 cells, the finding that overexpressed

HIF-1 α and HIF-2 α can bind without HIF-1 β raised the possibility that the endogenous HIF- α isoforms might be able to bind alone in other cellular settings. However, analysis of endogenous HIF ChIP-seq data from three further cell lines (HepG2, RCC4, HKC8) has not revealed any evidence to support widespread non-canonical binding of endogenous HIF- α without HIF-1 β .

Nevertheless, several possible caveats apply to this conclusion. Firstly, it is possible that HIF- α is able to bind without HIF-1 β at a very small minority of sites that might be missed by a statistical analysis of overall patterns. Indeed, HIF- α ChIP-seq signal in the absence of HIF-1 β signal was detected at a handful of sites in the different cell lines. However, with one exception the neighbouring genes were either not expressed or not regulated by hypoxia in RNA-seq analyses of these cell lines suggesting that they are either non-functional sites or false-positive HIF- α ChIP-seq signal. Either way, this does not alter the general overall pattern that was observed.

Secondly, although unlikely, it is possible that HIF- α is able to bind chromatin without HIF-1 β , but only at sites at which canonical HIF- α /HIF-1 β heterodimers also bind. This is a harder question to address. One might examine the ratio of HIF- α signal to HIF-1 β signal. However, whilst this clearly varies between sites, it is difficult to determine real signal from statistical noise. Ultimately, this possibility might be better addressed through further experimental studies on cell lines that have been engineered by CRISPR-Cas9 to delete HIF-1 β expression.

Finally, it is technically possible that HIF- α binds with an alternate dimerisation partner, but that this blocks binding of the antibody used in the ChIP experiments. Again this is unlikely, since both HIF- α antibodies used in these

experiments are polyclonal antibodies generated by large peptide immunogens and were generated to the C-terminal half of HIF- α that did not include the PAS domain usually responsible for dimerisation (Lau et al., 2007). Furthermore, non-canonical HIF- α ChIP-seq signal was observed in the overexpression studies, which used the same immunoprecipitating antibodies.

The lack of widespread HIF- α binding without HIF-1 β binding does not exclude the possibility that HIF- α might dimerise with paralogs of HIF-1 β such as ARNT2, ARNTL1 or ARNTL2. Through interrogation of our HKC8 and Hepg2 RNA-seq and RCC4 microarray data, expression levels of these HIF-1 β paralogs vary across the cell lines being used. However ARNTL2 is consistently expressed at similar levels to HIF-1 β , thus suggesting that ARNTL2 is a candidate alternative dimerisation partner of HIF- α . However, it is possible that, being paralogs, they might also be recognised by the HIF-1 β antibody used for the ChIP analysis, in which case they would generate apparent canonical signal.

Reciprocal analyses of HIF- α binding at HIF-1 β sites again indicated a lack of widespread HIF-1 β binding in the absence of HIF- α . Whilst the same caveats apply as above, it is also likely that other well described HIF-1 β dimerisation partners such as AhR, and SIM1 and SIM2 (Kewley et al., 2004) only bind HIF-1 β following appropriate stimulation of the cells to activate expression of the protein.

Taken together, these results indicate that HIF-1 α and HIF-2 α almost always bind chromatin in combination with HIF-1 β . Therefore, in future analyses I have included HIF-1 β signal in my definition of HIF-binding sites in order to give a greater degree of security to these sites.

Chapter 3: Identifying characteristic properties of *cis*-regulatory elements preferentially bound by either HIF-1 α or HIF-2 α

3.1 Introduction

Characteristics common to both HIF-1 α and HIF-2 α have been widely described in the literature. These include a common dimerisation partner (HIF-1 β) and a common mechanism of regulation through the PHD-VHL cellular oxygen sensing axis (Kaelin and Ratcliffe, 2008). Furthermore, HIF-1 α and HIF-2 α share remarkable structural similarity (Ema et al., 1997). Indeed, there is 83% shared amino acid sequence identity within the bHLH DNA binding domain of HIF-1 α and HIF-2 α ; 67% in the PAS domains, and 63% in the C-terminal domain. Thus, it is not surprising that each isoform binds to the same common consensus HRE motif on chromatin (Wenger et al., 2005).

However, despite this significant sequence similarity and shared consensus motif, HIF-1 (HIF-1 α /HIF-1 β heterodimer) and HIF-2 (HIF-2 α /HIF-1 β heterodimer) exhibit marked differences in chromatin binding across the genome, with HIF-1 tending to bind closer to the TSS (gene transcription start site) and HIF-2 binding more distal to gene promoters (Schodel et al., 2011). These distributional differences appear to be uninfluenced by the expression level of each isoform in overexpression studies (Salama et al., 2015). However, it is not clear how generalisable such differences in HIF-1 and HIF-2 pan-genomic binding are between different cell types.

HIF-1 and HIF-2 are known to regulate distinct sets of genes, and subsequently exert differing functional effects on a cell (Hu et al., 2003; Raval et al., 2005). In MCF7 cells, HIF-1 and HIF-2 have been shown to bind sites whereby the nearest gene tends to be upregulated in hypoxia (Schodel et al., 2011). The distinct nature of HIF-1 and HIF-2 binding sites and target genes is what determines their relative effects on cell behaviour.

Whilst more severe tumour hypoxia has been associated with a poorer prognosis in cancer patients (Semenza, 2003), separate modulation of HIF-1 α or HIF-2 α status can evoke a positive or negative effect on the progression of a tumour, depending on the cancer type (Keith et al., 2012). For instance, in clear cell renal carcinoma (ccRCC), expression of HIF-1 α has been shown to evoke a reduced capacity for the tumour to grow, whereas increased HIF-2 α expression has been shown to encourage tumour progression (Lidgren et al., 2005; Raval et al., 2005). Therefore, understanding what contributes to the differing scope of HIF-1 and HIF-2 binding sites in different cell types may subsequently enable identification of novel anti-cancer therapeutic strategies.

Meta-analysis of microarray experiments, obtained following hypoxic stimulation of a range of cell types, has indicated that there are a relatively small number of conserved hypoxia-regulated genes across cell types, with each cell type regulating its own private repertoire of genes in response to hypoxia (Ortiz-Barahona et al., 2010). The corollary of this is that HIF binding will also be highly cell-type specific, with relatively few common sites. Furthermore, since HIF-1 α is more widely expressed than HIF-2 α across different cell types (Kaelin and Ratcliffe, 2008) it is possible that each isoform exhibits different degrees of conservation in its binding. However, the extent to which HIF binding is conserved between cell types has been poorly studied to date. I will therefore implement analyses of ChIP-seq data to provide greater insight into this phenomenon.

In this chapter, I will explore the binding site characteristics of HIF-1 and HIF-2. Firstly, I will examine the conservation of HIF-1 and HIF-2 binding sites across cell types. I will then examine for trends in distance between binding sites

and the nearest gene, for differences in transcription factor motif enrichment, transcription factor binding itself and for differences in histone modifications enrichment among sites preferentially bound by either HIF-1 or HIF-2. Determining the characteristics that are specific to HIF-1 and HIF-2 preferentially bound sites will enable a better understand of the mechanisms that underpin the differences in HIF-1 and HIF-2 binding site selection.

3.2 Results

3.2.1 HIF binding site conservation

To help understand whether HIF-1 or HIF-2 binding sites exhibit differing degrees of conservation between different cell types, I examined for overlap between HIF-1 or HIF-2 binding sites in HepG2, RCC4 and HKC8 cells. I also utilised the previously published HIF-1 and HIF-2 ChIP-seq data from MCF7 cells (Schodel et al., 2011). In all cell lines assayed, ChIP-seq peaks were called using the MACS peak caller (Zhang et al., 2008b).

Whilst I acquired ChIP-seq data in HepG2, RCC4 and HKC8 cells in biological duplicate, the published MCF7 ChIP-seq datasets were only performed as single experiments. Where biological duplicate data was available, only ChIP-seq peaks that were identified in both biological replicates were considered further. Specifically, peaks were first called in each replicate independently. Peaks from replicate 1 that overlapped a peak in replicate 2 by one or more base pairs were then identified using BEDTools intersect (v2.26.0) (Dale et al., 2011). For a site to be termed a HIF-1 binding site in any of the four cell lines, a HIF-1 α peak that is present in both HIF-1 α ChIP-seq biological replicates must also be overlapping a HIF-1 β peak that is present in both HIF-1 β ChIP-seq biological replicates. Similar criteria were used to define HIF-2 binding sites.

Having defined HIF-1 and HIF-2 binding sites in each of the four cell lines, I then used the dba.overlap function in DiffBind (Ross-Innes et al., 2012) to establish the number of sites conserved between cell types, as well as using BEDTools intersect to identify specific sites that were unique or conserved across two or more of the four cell types. I have represented these data in the form of bar charts and Venn diagrams (Figure 1). Of the 2747 HIF-1 sites, 84 were

conserved across all four of the tested cell lines (Figure 1A). By comparison, 46 of the 4269 HIF-2 sites were conserved across all four cell lines (Figure 1B). Considering the proportion of sites conserved across two or more cell types, reveals that HIF-1 sites have a greater tendency to be conserved across cell types than HIF-2 sites (Figure 1C). Specifically, 75% of HIF-1 sites are unique to a single cell line, whereas 85% of HIF-2 sites are unique to a single cell line, whilst 3.1% of HIF-1 sites and 1.1% of HIF-2 sites are conserved across all four cell lines tested.

It is important to recognise that there are a number of datasets that are included in our analyses that one might regard as having a disproportionately high number number binding sites. This represents a caveat to the analysis. Namely, there are significantly greater HIF-1 binding sites in HepG2 cells than in RCC4, HKC8 and MCF7 cells. Likewise, there are significantly greater HIF-2 binding sites in HKC8 cells than in HepG2, RCC4 and MCF7 cells. Given that each dataset is comprised of high confidence peaks, the disproportionately high number of peaks in the aforementioned HIF-1 and HIF-2 datasets are likely to be a true biological phenomenon that may reflect the relative expression levels of HIF-1 and HIF-2 protein in these cells. Including these high volume datasets in the four-way analysis is inevitably going to lead to significant proportion of the peaks within these high-volume datasets not being conserved, and thus leading to a diminishment of the overall binding site conservation percentage.

Analyses performed in Chapter 2, illustrated the limitations of using ChIP-seq peak callers to qualitatively define overlapping and non-overlapping sites in a binary fashion. Whilst these methodologies tend to underestimate the degree of overlap between two ChIP-seq datasets, in the current analysis the same

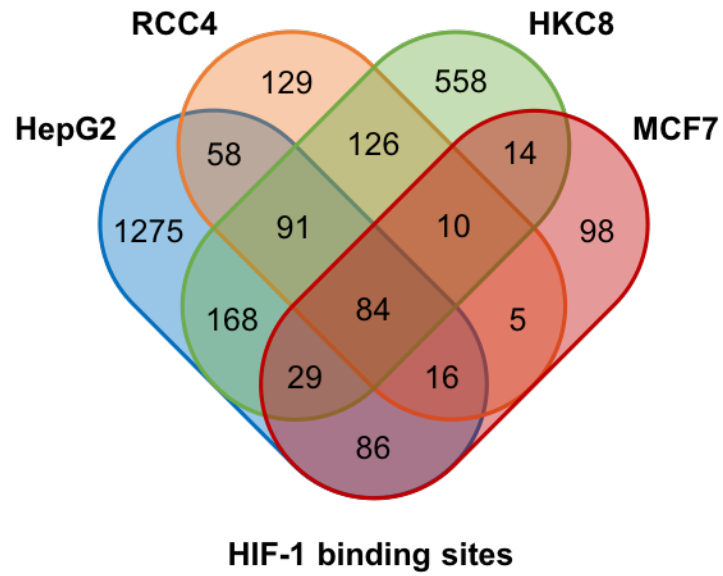
pipelines were applied to both HIF-1 and HIF-2, so comparison of conservation between the two isoforms is valid. Nevertheless, to increase confidence in these results, I have also examined quantitative ChIP-seq signal intensity in all four cell lines at all sites identified by the MACS peak caller as binding either HIF-1 or HIF-2 in HKC8 cells.

I first determined the number of HIF-1 α reads in HKC8 cells at HIF-1 peak summits (peak summits are 1 bp coordinates within ChIP-seq peaks whereby the ChIP-seq signal for the factor being subjected to ChIP is at its highest) that had previously been defined as binding HIF-1 in HKC8 cells-only. Read coverage was determined using the BEDTools multicov function. Normalised read coverage (RPKM) may also be defined as the 'peak height' (i.e. the number of overlapping reads at the point of maximum signal). ngs.plot (v2.61) (Shen et al., 2014) was used to map HIF-1 α read counts per million (CPM) in HepG2, RCC4, HKC8 and MCF7 cells within the sorted HIF-1 summits (Figure 2). In each of the four heatmaps, the HIF-1 peak summits that were deemed to be bound specifically in HKC8 cells, were sorted by HIF-1 α peak height in the corresponding cell type. Identical bioinformatics protocols were applied to HIF-2 peak summits deemed to be bound specifically in HKC8 cells. A small degree of HIF-1 α and HIF-2 α signal was observed in HepG2, RCC4 and MCF7 cells (as represented by red colour intensity in the heatmaps), whereas signal was strong in HKC8 cells. Based upon the heatmap data, it is likely that only a small proportion of the sites specifically bound in HKC8 cells by HIF-1 and HIF-2 represent false-negative HIF-1 α or HIF-2 α peaks in the other three cell lines. Therefore, these data largely support the binding site conservation data illustrated by Venn diagrams earlier in this chapter.

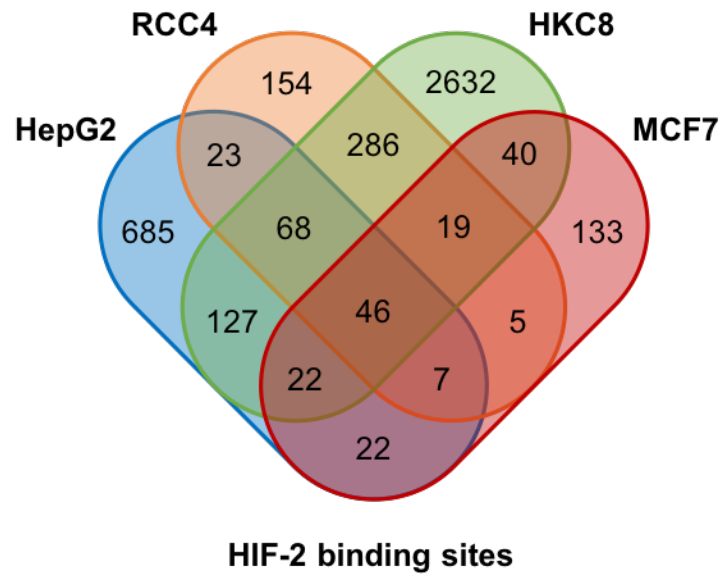
A region upstream of the *SEMA4B* gene illustrates a binding site that is conserved across all cell lines tested (Figure 2B). Conversely, intronic regions of the *KANK2* gene and the *CROCC* gene are good examples of sites bound specifically in HKC8 cells by HIF-1 or HIF-2 respectively (Figure 2C-D).

I next examined how the degree of binding site conservation between cell types was reflected in the distribution of binding around the TSS of the nearest genes. Owing to the greater evolutionary conservation of promoter-proximal *cis*-regulatory elements relative to promoter-distal *cis*-elements (Long et al., 2016), I hypothesised that the more conserved binding sites would distribute closer to the nearest gene. I therefore categorised HIF-1 and HIF-2 binding sites according to the number of cell lines in which they were present. The distance to the nearest TSS was then calculated and the frequency distribution plotted as a histogram for each set of sites (Figure 3). In keeping with previous results (Schodel et al., 2011), HIF-1 sites were more promoter-proximal than HIF-2 sites for all degrees of conservation between cell lines. In addition, both HIF-1 sites and HIF-2 sites were progressively more promoter-proximal as the degree of conservation between cell lines increased. Thus conserved sites tend to bind close to the gene promoter, whilst cell-type specific sites bind at greater distances from the nearest gene.

A



B



C

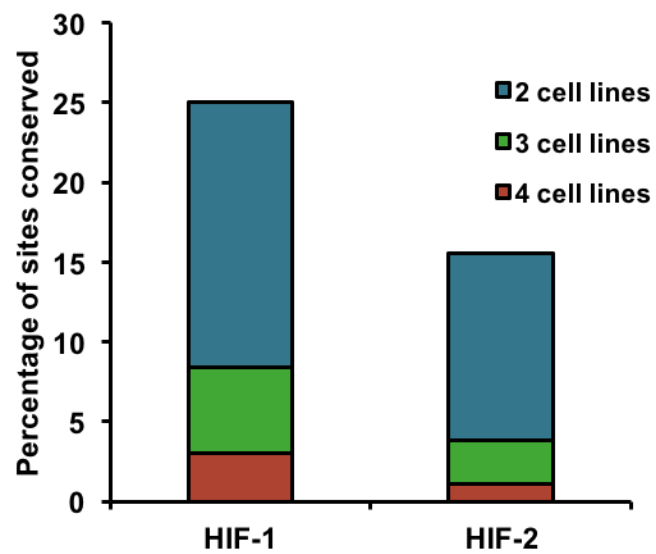


Figure 1. Conservation of HIF binding sites across cell lines.

(A) Venn diagram illustrating the overlap between HIF-1 binding sites across four cell lines: HepG2, RCC4, HKC8 and MCF7.

(B) Venn diagram illustrating the overlap between HIF-2 binding sites across four cell lines: HepG2, RCC4, HKC8 and MCF7.

(C) Bar chart depicting the percentage of HIF-1 and HIF-2 binding sites conserved in two, three or four cell lines.

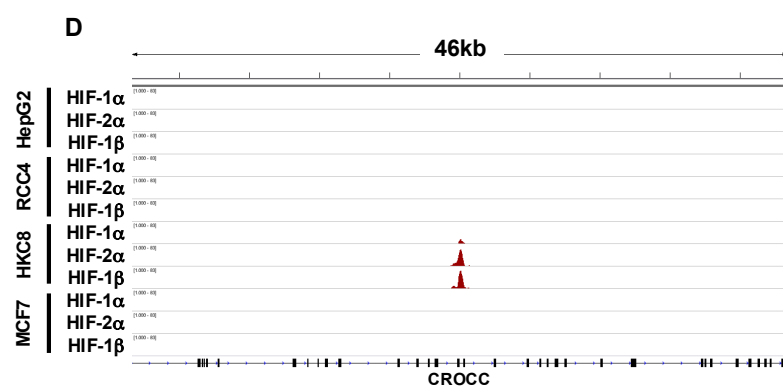
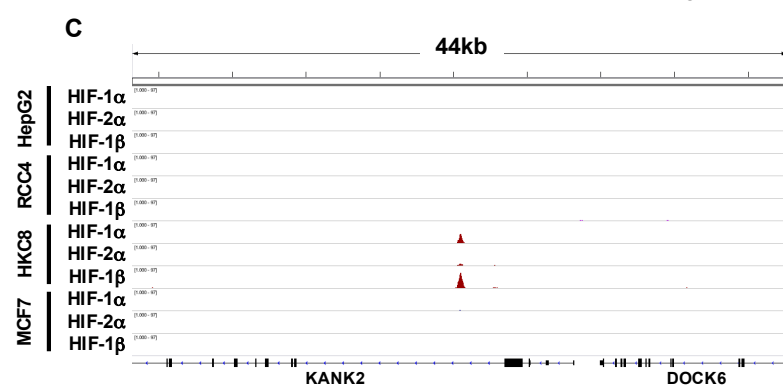
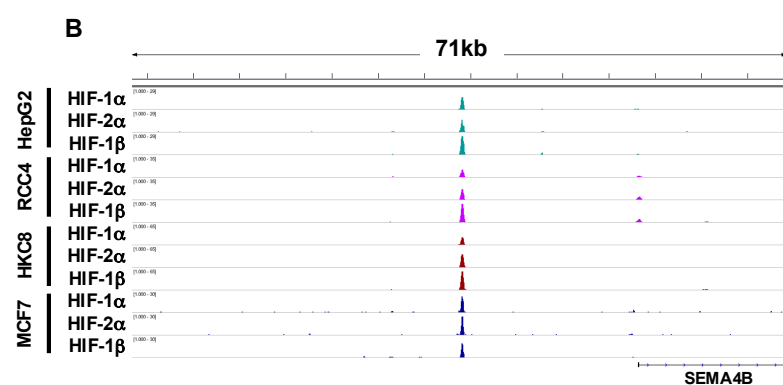
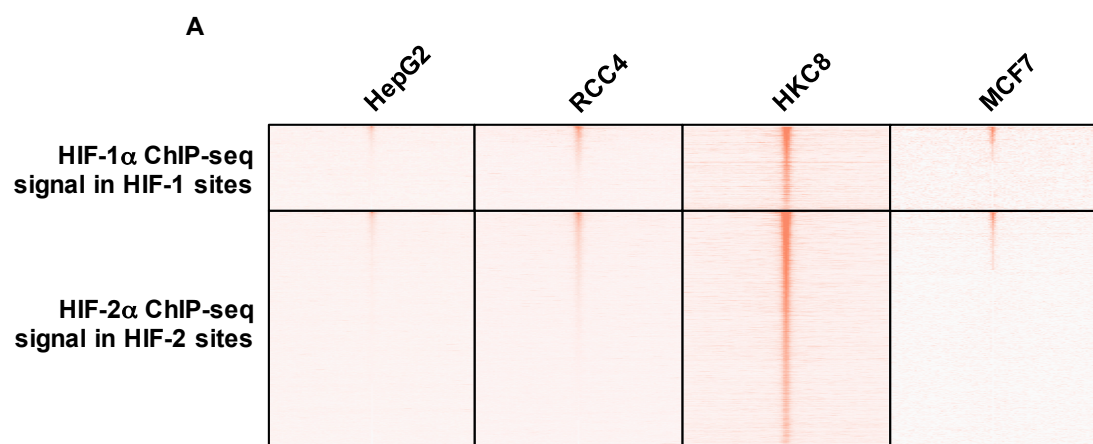


Figure 2. Corroboration of peak caller binding conservation data via assessment of ChIP-seq signal intensity.

(A) Heatmaps showing HIF-1 α and HIF-2 α ChIP-seq signal (CPM as colour intensity) in HepG2, RCC4, HKC8 and MCF7 cells for regions ± 5 kb around HIF-1 and HIF-2 peak summits in HKC8 cells (x-axis). All heatmaps are sorted based upon the HIF-1 α or HIF-2 α peak height in the corresponding cell type (vertical axis). ChIP-seq experiments were performed in biological duplicate on HepG2, RCC4 and HKC8 cells, whereas MCF7 ChIP-seq experiments were performed in singlicate. HepG2, HKC8 and MCF7 cells were incubated for 16 hours in 0.5% O₂ prior to harvesting, whereas RCC4 cells were harvested in normoxia. HIF- α peaks were only considered if called in both independent biological replicates. Merged .bam files were obtained by combining two biological replicates. These merged .bam files were used to map ChIP-seq signal intensity (RPM) across the defined regions depicted by the heatmaps.

(B) A 71kb genomic region around the *SEMA4B* gene depicting representative HIF-1 α , HIF-2 α and HIF-1 β ChIP-seq signals in HepG2 (green), RCC4 (pink), HKC8 (red) and MCF7 (blue) cells.

(C) A 44kb genomic region around the *KANK2* gene depicting representative HIF-1 α , HIF-2 α and HIF-1 β ChIP-seq signals in HepG2 (green), RCC4 (pink), HKC8 (red) and MCF7 (blue) cells.

(D) A 46kb genomic region around the *CROCC* gene depicting representative HIF-1 α , HIF-2 α and HIF-1 β ChIP-seq signals in HepG2 (green), RCC4 (pink), HKC8 (red) and MCF7 (blue) cells.

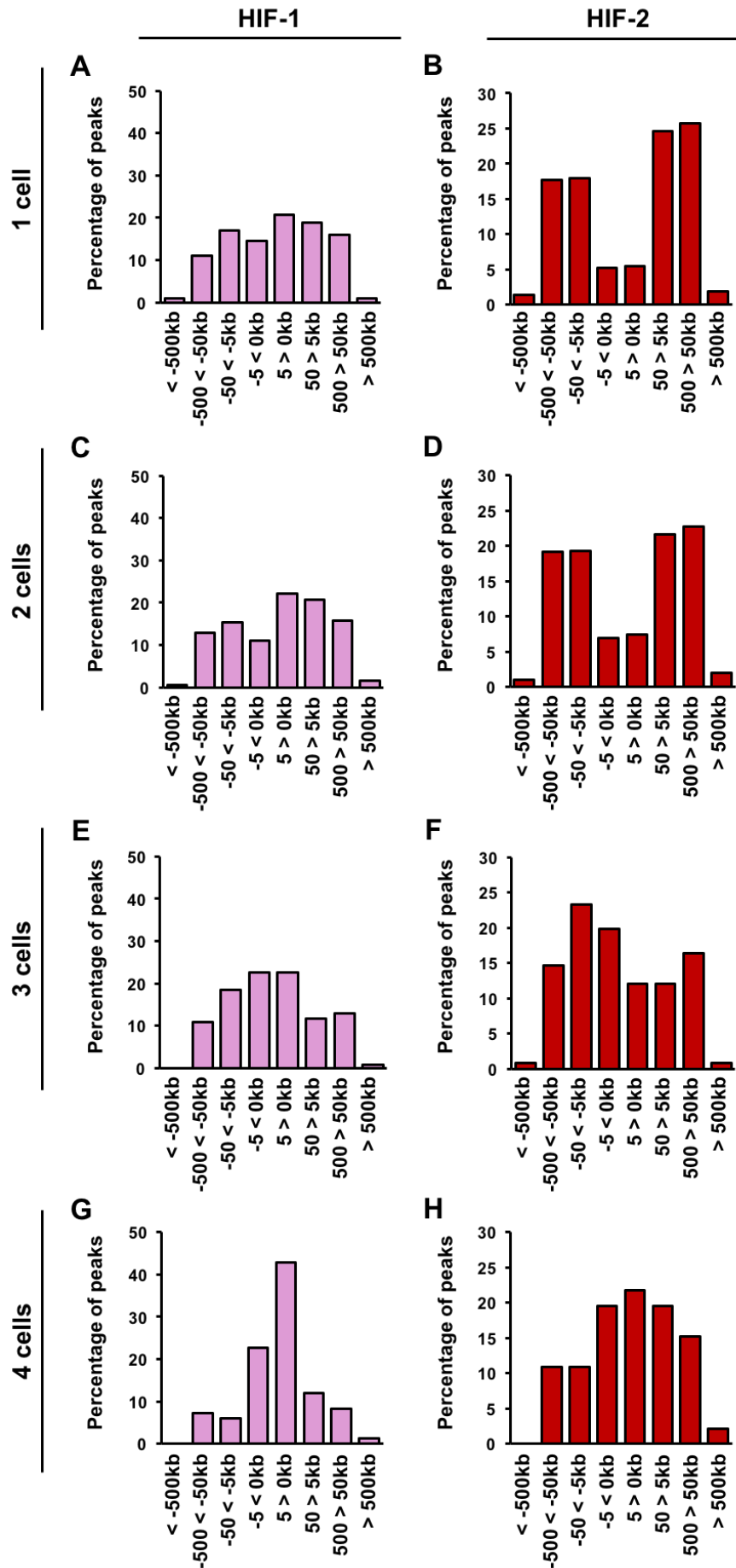


Figure 3. Distribution of HIF binding sites relative to the nearest TSS, according to their conservation between cell lines.

(A) Histogram depicting the proportion of HIF-1 binding sites present in only one cell line that fall within the defined bins based upon the distance between the binding site and the nearest TSS.

(B) Histogram as in (A), for HIF-2 sites present in only one cell line.

(C) Histogram as in (A), for HIF-1 sites conserved in two cell lines.

(D) Histogram as in (A), for HIF-2 sites conserved in two cell lines

(E) Histogram as in (A), for HIF-1 sites conserved in three cell lines.

(F) Histogram as in (A), for HIF-2 sites conserved in three cell lines

(G) Histogram as in (A), for HIF-1 sites conserved in four cell lines.

(H) Histogram as in (A), for HIF-2 sites conserved in four cell lines.

3.2.2 Stratification and characterisation of sites preferentially bound by either HIF-1 or HIF-2

Prior studies of HIF binding have identified pan-genomic HIF-1 and HIF-2 binding sites in a variety of cell settings (Mole et al., 2009; Xia and Kung, 2009; Tanimoto et al., 2010; Schodel et al., 2011; Villar et al., 2012; Schodel et al., 2012; Choudhry et al., 2014; Tausendschön et al., 2015; Salama et al., 2015; Grampp et al., 2016; Platt et al., 2016). These sets of HIF-1 and HIF-2 binding sites contained significant numbers of shared sites as well as varying numbers of HIF-1 or HIF-2 preferentially-bound sites.

To better understand the characteristics that determine a HIF-1 vs. a HIF-2 binding site, I first used a quantitative approach to stratify sites preferentially bound by either HIF-1 or HIF-2 and exclude sites that bound both isoforms equally. Having previously established in Chapter 2 that the vast majority of endogenous HIF- α sites are concurrently bound by HIF-1 β , I first identified all HIF-1 β ChIP-seq peaks in HepG2, RCC4 and HKC8 cell lines using the ChIP-seq peak caller, MACS as before. Only the HIF-1 β peaks identified in both biological replicates were considered for further analysis. Specifically, peaks were first called in each replicate independently. Peaks from replicate 1 that overlapped a peak in replicate 2 by one or more base pairs were identified using BEDTools intersect, then subsequently maintained for further analysis.

I then stratified sites preferentially bound by HIF-1 or HIF-2 by sorting HIF-1 β peaks according to the ratio of HIF-1 α and HIF-2 α ChIP-seq signal (mean RPKM, read counts per million mapped reads per kilobase, across the two biological replicates) within the HIF-1 β sites. HIF-1 α (blue) and HIF-2 α (red) RPKM moving average within the sorted HepG2, RCC4 and HKC8 HIF-1 β peaks

were plotted in the form of a line plot (Figure 4A-C). Sites were defined as preferentially bound by HIF-1 or HIF-2 when the ratio of HIF-1 α to HIF-2 α RPKM was 1.5-fold greater than that of the other HIF- α subunit.

In order to validate the functionality of sites preferentially bound by either HIF-1 or HIF-2, sites in HepG2 cells were subjected to gene set enrichment analysis (GSEA) to associate the binding sites with hypoxia regulation of the neighbouring genes (Subramanian et al., 2005). Total RNA was harvested from HepG2 cells that had been cultured in normoxia or in 0.5% O₂ for 16 hours. RNA was harvested in biological triplicate for each condition and next generation sequencing libraries were prepared from RNA by polyA⁺ selection. DESeq2 analysis was performed to determine the normoxic vs. hypoxic fold change and significance of differential expression across all detected transcripts. Genes were ranked by a combined metric, incorporating the fold change and significance of differential expression as determined by DESeq2 (Xiao et al., 2014). Sites preferentially bound by either HIF-1 or HIF-2 were annotated by their nearest protein coding gene, and these genes were subsequently assessed for enrichment among the ranked RNA-seq gene list.

GSEA indicated that genes closest to the HIF-1 ($p < 0.001$) and HIF-2 ($p = 0.005$) preferentially bound sites were significantly enriched amongst hypoxia upregulated genes, therefore validating the functionality of these sites (Figure 5A-B). Representative loci preferentially bound by HIF-1 and HIF-2, respectively, have been illustrated by ChIP-seq tracks (Figure 5C-D).

Two prior pan-genomic HIF binding studies published by the Ratcliffe Group have shown that HIF-1 and HIF-2 binding sites have distinct distributions relative to the nearest gene to each binding site (Schodel et al., 2011; Salama et

al., 2015). HIF-1 binding sites in MCF7 cells were shown to exhibit a more promoter-proximal distribution relative to HIF-2, whose binding sites were more promoter-distal. And similarly, in 786O cells overexpressing either HIF-1 α or HIF-2 α , HIF-1 α binding sites were more promoter-proximal than HIF-2 α sites. I therefore wished to assess whether regions preferentially bound by either endogenous HIF-1 or HIF-2 adopt the same distributions with respect to the nearest gene across a wider array of cell lines. For each set of HIF-1 and HIF-2 binding sites, the distance from the HIF-binding site to the nearest TSS was calculated and the frequency distribution then plotted as a histogram (Figure 6).

In HepG2 cells, 65.1% of preferentially bound HIF-1 sites reside less than 5kb from the TSS of the nearest gene, whereas only 10.2% of preferentially bound HIF-2 sites within these regions (Figure 6A-B). Conversely, only 11.2% of preferentially bound HIF-1 sites in HepG2 cells lie more than 50 kb from the TSS of the nearest gene, whereas 52.7% of preferentially bound HIF-2 sites lie more than 50 kb from the nearest gene. This trend is similarly conserved across all three of the cell lines in which endogenous HIF- α distributions were assayed (Figure 6C-F).

Although HIF-1 and HIF-2 have different distributions around the nearest TSS, it is possible that only one isoform is being selectively distributed and that the other simply occupies accessible HRE motifs in a random fashion. I therefore determined the distance between each HRE situated in accessible chromatin relative to the nearest TSS and plotted the frequency distribution as a histogram (Figure 6G). Accessible HREs were determined by intersecting all genomic RCGTG sequences with DNaseI hypersensitive regions, as determined by the ENCODE Project in normoxic HepG2 cells (www.encodeproject.org; accession

number in Appendix 8.1), using BEDTools intersect. 43% of accessible HREs are situated less than 5 kb from the TSS of the nearest gene. Meanwhile 22.4% of accessible HREs are situated more than 50 kb from the TSS of the nearest gene. The disparity between the distribution of accessible HREs with respect to sites preferentially bound by both HIF-1 and HIF-2 suggests that neither HIF-1 nor HIF-2 randomly select the HREs to which they bind.

The relative distributions of HIF-1 and HIF-2 appear to be fixed regardless of cell type. Of further note, neither HIF-1 nor HIF-2 appear to adopt a pan-genomic distribution that could be explained by random selection of accessible HREs. I therefore postulated that there may be distinct underlying DNA sequence determinants unique to both sets of preferentially bound sites that might be directing HIF-1 and HIF-2 to their respective binding sites. Any identified sequence specificities may be further indicative of transcription factors that HIF-1 or HIF-2 may be cooperating with, thus forming part of their relative binding site selection processes.

To enable motif analysis of sites preferentially bound by each isoform, the coordinates of HIF-1 and HIF-2 peak summits within the respective preferentially bound sites were extended by ± 150 bp. Repetitive sequences within these regions were then masked using RepeatMasker (v4.0.3) (www.repeatmasker.org). Enriched sequence motifs were identified using AME (v4.11.4) (McLeay and Bailey, 2010), in which a library of 519 known motifs in vertebrates (JASPAR Vertebrates 2016) were each tested for statistically significant enrichment relative to a set of randomly shuffled control regions. For each set of test regions, the complete list of motifs enriched with an adjusted *p*-value of less than 0.05, in addition to their respective JASPAR Database IDs, are

given in tables in Appendix 8.5. The top four motifs enriched for each set of binding sites are illustrated in Figures 7-9.

Published data HIF-1 and HIF-2 ChIP-seq from MCF7 cells has previously confirmed the binding of both HIF-1 and HIF-2 to the hypoxia response element (Wenger et al., 2005; Schodel et al., 2011). This motif (MA0259.1) was the most enriched motif at all sites preferentially bound by HIF-1 or HIF-2 in HepG2, RCC4 and HKC8 cells (Figures 7-9).

The SP1 (MA0079.3) motif was the second most enriched motif at HIF-1 sites in HepG2 cells (adj. $p = 4.62\text{e-}79$) (Figure 7), whereas the FOXKD2 (MA0847.1) motif was the second most enriched motif at HIF-2 sites (adj. $p = 6.6\text{e-}27$) (Figure 7B). GC-rich motifs tended to be abundantly enriched in preferentially bound HIF-1 sites in HepG2 cells (Figure 7). These included the E-box and further variants on the SP1/2 and KLF zinc-finger motifs. HIF-2 preferentially bound sites showed a strong enrichment for the forkhead box transcription factor motifs (Figure 7).

In RCC4 cells, the ZNF263 (MA0529.1) motif was the second most enriched motif in preferentially bound HIF-1 sites (adj. $p = 2.5\text{e-}17$) (Figure 8), whereas the AP-1 (MA0489.1) motif was the second most enriched motif in HIF-2 sites (adj. $p = 1.18\text{e-}08$) (Figure 8). In a similar vein to HepG2 cells, GC-rich motifs tended to be abundantly enriched in preferentially bound HIF-1 sites in RCC4 cells (Figure 8), including the E-box and further variants on the SP1/2 and KLF zinc-finger motifs. However, unlike HepG2 cells, HIF-2 preferentially bound sites in RCC4 cells showed a strong enrichment for multiple variants of the AP-1 transcription factor motif (Figure 8).

In HKC8 cells, the HEY2 (MA0649.1) E-box motif was the second most enriched motif in preferentially bound HIF-1 sites (adj. $p = 6.58\text{e-}15$) (Figure 9), whereas the AP-1 (MA0489.1) motif was the second most enriched motif in HIF-2 sites (adj. $p = 1.34\text{e-}118$) (Figure 9). Like both HepG2 and RCC4 cells, GC-rich motifs are abundantly enriched in preferentially bound HIF-1 sites (Figure 9), including the E-box and further variants on the SP1/2 and KLF zinc-finger motifs. Like in RCC4 cells, HIF-2 preferentially bound sites in HKC8 cells showed a strong enrichment for multiple variants of the AP-1 transcription factor motif (Figure 9).

A recognisable trend is that GC-rich motifs tend to be enriched within sites preferentially bound by HIF-1, and to a much lesser extent within HIF-2 sites. The AP-1 motif tends to be strongly enriched in sites preferentially bound by HIF-2. This is particularly pronounced in the renal cells tested (HKC8 and RCC4). However, this enrichment is less significant in HIF-2 sites in HepG2 cells. In HepG2 cells there appears to be strong cell-specific enrichment of the forkhead motif.

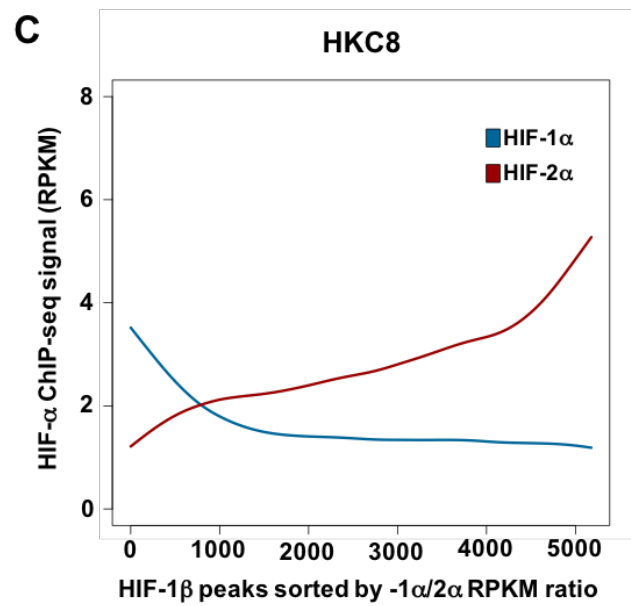
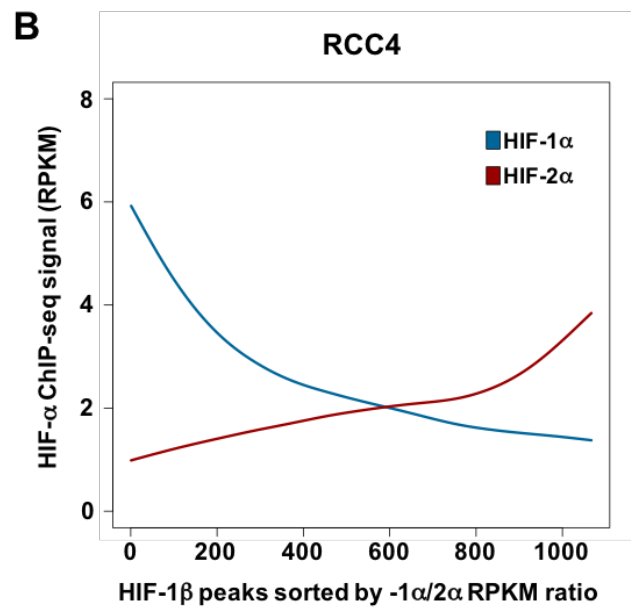
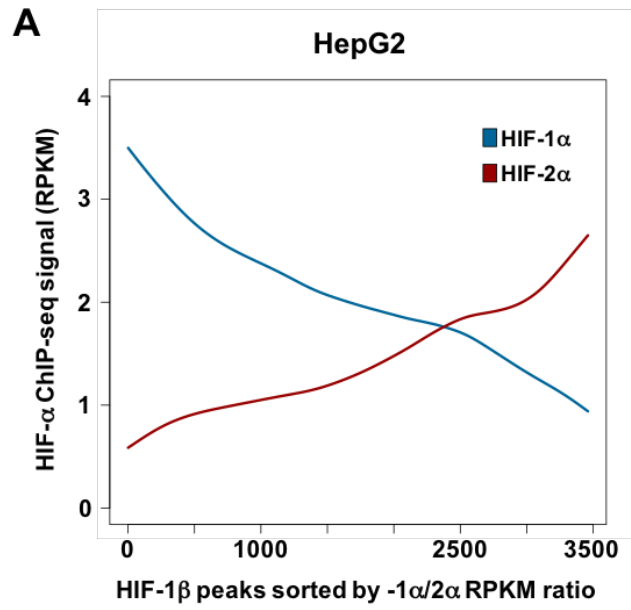


Figure 4. Identification of sites preferentially bound by HIF-1 or HIF-2.

(A) Line plot showing the moving average of HIF-1 α (blue) and HIF-2 α (red) RPKM within the totality HIF-1 β sites in HepG2 cells that have been sorted by the ratio of HIF-1 α to HIF-2 α RPKM. Sites deemed to be preferentially bound by one HIF- α isoform or the other when their RPKM is greater than 1.5x the RPKM of the other isoform.

(B) Line plot as in (A) but in RCC4 cells.

(C) Line plot as in (A) but in HKC8 cells.

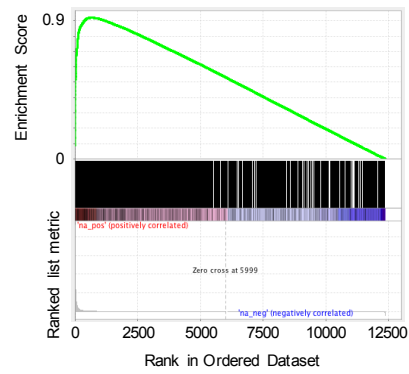
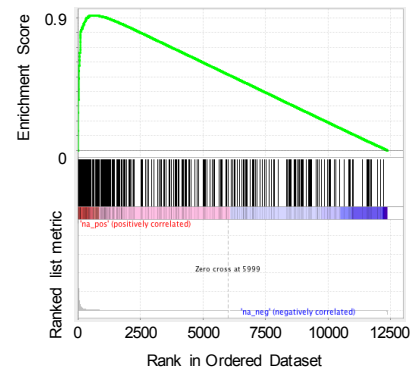
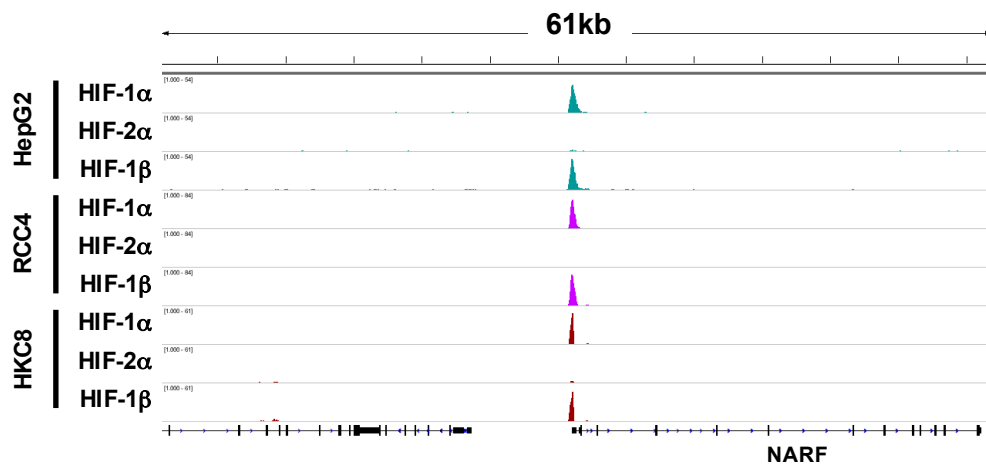
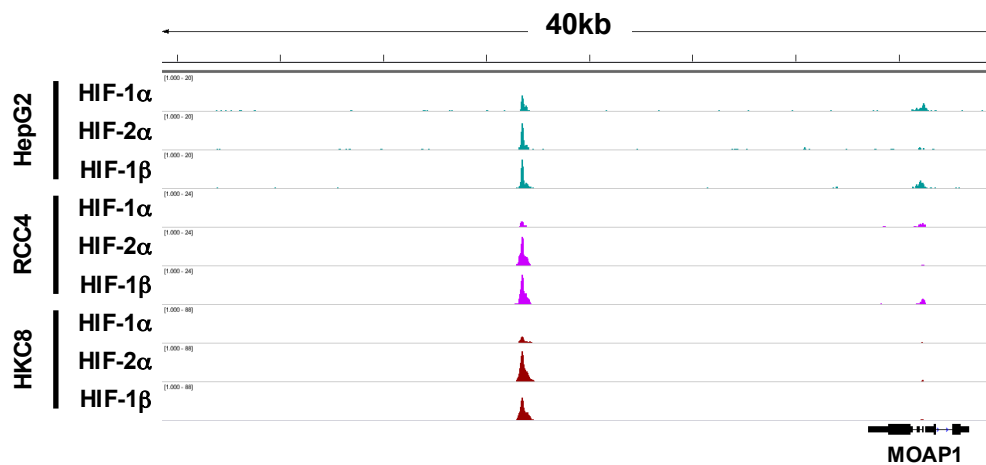
A**B****B****C**

Figure 5. Functional association of HIF isoform preferentially bound sites.

(A) Association of sites preferentially bound by HIF-1 in HepG2 cells with hypoxia regulated genes via gene set enrichment analysis (GSEA) ($p < 0.001$).

(B) Association of sites preferentially bound by HIF-2 in HepG2 cells with hypoxia regulated genes via GSEA ($p = 0.005$).

(C) A 61 kb genomic region around the *NARF* gene illustrating a site preferentially bound by HIF-1. Representative HIF-1 α , HIF-2 α and HIF-1 β ChIP-seq signals in HepG2 (green), RCC4 (pink) and HKC8 (red) cells.

(D) A 40 kb genomic region around the *MOAP1* gene illustrating a site preferentially bound by HIF-2. Representative HIF-1 α , HIF-2 α and HIF-1 β ChIP-seq signals in HepG2 (green), RCC4 (pink) and HKC8 (red) cells.

Figure 6. Distribution of HIF isoform preferentially bound sites relative to the nearest TSS.

(A) Frequency distribution histogram depicting the proportion of preferentially bound HIF-1 sites in HepG2 cells based upon the distance between a site and the TSS of the nearest gene.

(B) Histogram as in (A), however HIF-2 sites plotted.

(C) Histogram as in (A), however HIF-1 sites in RCC4 cells plotted.

(D) Histogram as in (A), however HIF-2 sites in RCC4 cells plotted.

(E) Histogram as in (A), however HIF-1 sites in HKC8 cells plotted.

(F) Histogram as in (A), however HIF-2 sites in HKC8 cells plotted.

(G) Histogram as in (A), however accessible HREs in HepG2 cells plotted.

HIF-1 sites

HIF-2 sites

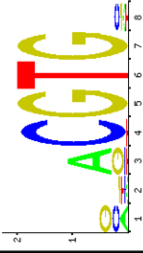
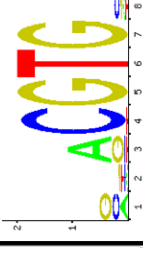
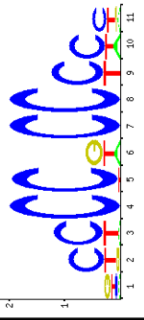
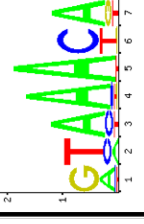
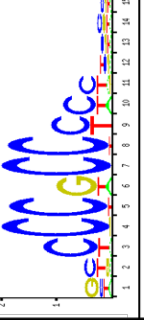
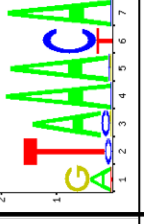
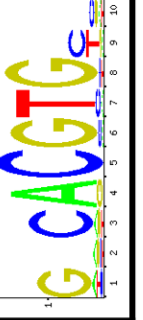
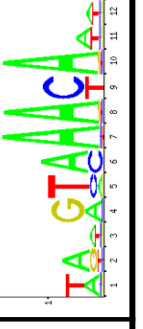
Motif	Transcription factor	Database ID	Adjusted <i>p</i> -value	Motif	Transcription factor	Database ID	Adjusted <i>p</i> -value
	ARNT::HIF1A	MA0259.1	2.32E-197		ARNT::HIF1A	MA0259.1	1.94E-40
	SP1	MA0079.3	4.62E-79		FOXD2	MA0847.1	6.60E-27
	SP2	MA0516.1	8.78E-76		FOXL1	MA0033.2	1.05E-25
	HEY2	MA0649.1	7.02E-73		FOXC2	MA0846.1	1.19E-24

Figure 7. Motif analysis of sites preferentially bound by either HIF-1 or HIF-2 in HepG2 cells.

Peak summits of the respective HIF-1 and HIF-2 preferentially bound sites in HepG2 cells were determined and 150 bp were subsequently added 5' and 3' to the summit coordinates. The JASPAR (2016) Vertebrates library of known motifs were then tested for enrichment within these sites relative to shuffled input sequences. Diagrammatic representation of the four most enriched motifs are shown, in addition to the JASPAR Database ID and adjusted *p*-value.

HIF-1 sites

HIF-2 sites

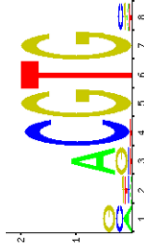
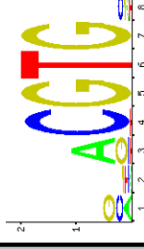
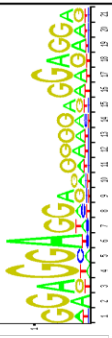
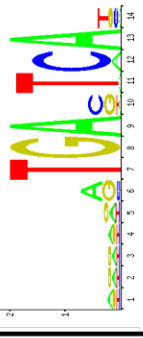
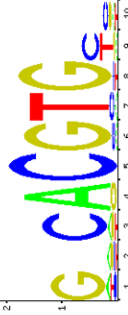
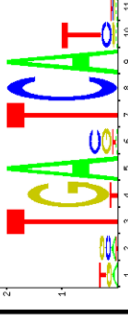
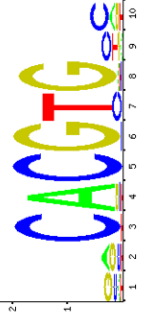
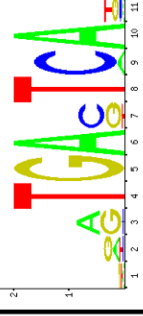
Motif	Transcription factor	Database ID	Adjusted <i>p</i> -value	Motif	Transcription factor	Database ID	Adjusted <i>p</i> -value
	ARNT::HIF1A	MA0259.1	2.22E-45		ARNT::HIF1A	MA0259.1	8.31E-26
	ZNF263	MA0528.1	2.85E-17		JUN(var.2)	MA0489.1	1.18E-08
	HEY2	MA0649.1	3.94E-17		FOS	MA0476.1	6.29E-08
	HEY1	MA0823.1	1.58E-14		JUNB	MA0490.1	1.33E-07

Figure 8. Motif analysis of sites preferentially bound by either HIF-1 or HIF-2 in RCC4 cells.

Peak summits of the respective HIF-1 and HIF-2 preferentially bound sites in RCC4 cells were determined and 150 bp were subsequently added 5' and 3' to the summit coordinates. The JASPAR (2016) Vertebrates library of known motifs were then tested for enrichment within these sites relative to shuffled input sequences. Diagrammatic representation of the four most enriched motifs are shown, in addition to the JASPAR Database ID and adjusted *p*-value.

HIF-1 sites

HIF-2 sites

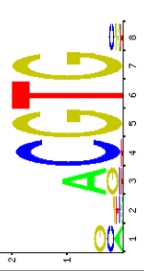
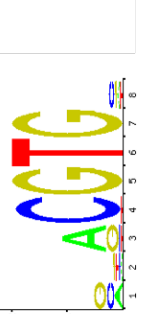
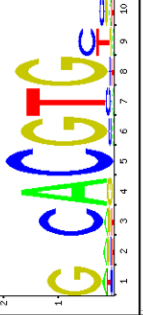
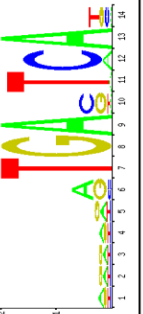
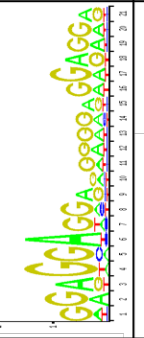
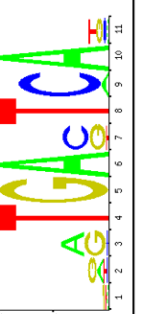
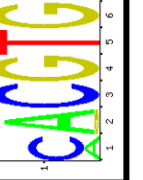
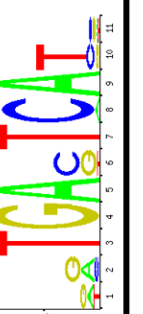
Motif	Transcription factor	Database ID	Adjusted p-value	Motif	Transcription factor	Database ID	Adjusted p-value
	ARNT::HIF1A	MA0259.1	2.22E-45		ARNT::HIF1A	MA0259.1	1.09E-196
	HEY2	MA0649.1	6.85E-15		JUN(var.2)	MA0489.1	1.34E-118
	ZNF263	MA0528.1	1.12E-13		JUNB	MA0490.1	9.77E-116
	ARNT	MA0004.1	1.26E-13		JUND	MA0491.1	4.05E-113

Figure 9. Motif analysis of sites preferentially bound by either HIF-1 or HIF-2 in HKC8 cells.

Peak summits of the respective HIF-1 and HIF-2 preferentially bound sites in HKC8 cells were determined and 150 bp were subsequently added 5' and 3' to the summit coordinates. The JASPAR (2016) Vertebrates library of known motifs were then tested for enrichment within these sites relative to shuffled input sequences. Diagrammatic representation of the four most enriched motifs are shown, in addition to the JASPAR Database ID and adjusted *p*-value.

3.2.3 Association of sites preferentially bound by HIF-1 or HIF-2 with ENCODE transcription factor ChIP-seq data

The association of different transcription factor binding motifs with sites that preferentially bind either HIF-1 or HIF-2 suggests that these factors may be contributing to the different binding distributions observed for each isoform. Furthermore, the different associations observed in different cell lines suggests that they may also play a role in helping to define cell-specific patterns of HIF binding. However, interpreting motif enrichment has a number of caveats. Firstly, the presence of a motif does not necessarily mean that the transcription factor will be bound. As an example there are over 2×10^6 HREs within the genome, however HIF only binds to a subset of these, even within accessible regions of chromatin (Schodel et al., 2013). Secondly, multiple transcription factors may be capable of binding the same or a closely related motif, so it may not be possible to know which of these transcription factors is actually binding. Thirdly, an extension of this is that transcription factors that bind to motifs that share homology with the HRE motif (e.g. an E-box, CACGTG) may be enriched as a consequence of enriching for the HRE-motif and not representing co-incident binding at all. Finally, GC content varies throughout the genome and is generally higher close to gene promoters. Thus GC-rich motifs would be expected to be enriched at HIF-1 binding sites as compared to HIF-2 binding sites given that HIF-1 binds more frequently close to the promoter. It should be possible to control for this using carefully matched control regions. However, it is also possible that factors that bind these motifs are the reason why HIF-1 binds more close to the promoter more frequently than HIF-2.

In an attempt to circumvent some of these issues, I examined the coincidence between sites preferentially bound by HIF-1 or by HIF-2 and transcription factor binding signal as defined by ChIP-seq analyses. Rather than performing ChIP-seq analysis of multiple transcription factors myself, I made use of the many ChIP-seq datasets published by the ENCODE (Encyclopedia of DNA Elements) Consortium for normoxic HepG2 cells and compared these with my ChIP-seq analysis of HIF binding in the same cell line. There are datasets from over 80 transcription factor ChIP-seq experiments publicly available, as well as ChIP-seq datasets of histone modifications, and DNase-seq analysis of chromatin accessibility (www.encodeproject.org; Accession numbers for all ENCODE datasets are listed in Appendix 8.1 and 8.2). The ENCODE datasets were derived from normoxic culture. Whilst it is possible that some of these transcription factors may be affected by oxygen levels, I argued that associations drawn between these factors and HIF should provide a good first approximation to their binding in hypoxia.

I firstly determined which transcription factor binding sites overlapped sites that preferentially bound either HIF-1 (1749 sites) or HIF-2 (476 sites). I downloaded all available ENCODE .bed files listing the peak coordinates for each transcription factor in HepG2 cells, as generated by the ENCODE ChIP-seq analysis pipeline. Binding sites for each transcription factor were intersected in turn with sites that preferentially bound either HIF-1 or HIF-2 using BEDTools intersect (v2.26.0). An overlapping peak was defined as any region that overlapped a HIF site by one or more base pairs. The statistical significance for the number of overlapping sites compared to that expected by random chance was then determined using a hypergeometric test for each transcription factor.

The transcription factors with binding sites that overlap HIF-1 preferentially bound sites with a $p < 1.0\text{e-}30$ are illustrated by a bar chart (blue), with transcription factors sorted by statistical significance (Figure 10A). Also illustrated is the number of peaks for each transcription factor, and the number of those peaks that are overlapping with a HIF-1 preferentially bound site. MAX interactor 1 (MXI1) peaks overlapped with 653 HIF-1 sites, and this was the most significantly overlapping transcription factor ($p = 4.67\text{e-}193$).

Similarly, the transcription factors with binding sites that overlap HIF-2 preferentially bound sites with a $p < 1.0\text{e-}30$ are illustrated by a bar chart (red), with transcription factors again sorted by statistical significance (Figure 10B). Transcription factor 12 (TCF12) sites overlapped 119 HIF-2 sites, and this was the most significant HIF-2 overlapping transcription factor (red) ($p = 5.23\text{e-}112$; Figure 10B).

Finally, the entire set of HIF-1 β sites (3460) were intersected with the ENCODE transcription factors (Figure 10C, purple bars) and MXI1 was again the most overlapping factor ($p = 2.68\text{e-}226$). 1079 HIF-1 β sites overlapped MXI1 peaks.

To corroborate these observations, I also generated heatmaps to show quantitative binding differences for these transcription factors at HIF-1 and HIF-2 preferentially bound sites (Figure 11). HIF-1 and HIF-2 peak summits were sorted by either HIF-1 α or HIF-2 α ChIP-seq signal (RPKM) respectively. ngs.plot (v2.61) was used as before to plot transcription factor ChIP-seq signal (read counts per million, CPM) against distance either side of the sorted HIF-1 and HIF-2 peak summits.

As expected, HIF-1 α signal was greater at HIF-1 peak summits than HIF-2 α , whilst HIF-2 α signal was greater at HIF-2 peak summits. HIF-1 β signal was equal at both HIF-1 α and HIF-2 α peak summits. GABPA and YY1 are two transcription factors that are enriched at HIF-1, but not HIF-2, binding sites. This finding is corroborated by the heatmaps, which show greater GABPA and YY1 ChIP-seq signal within regions preferentially bound by HIF-1 than in regions preferentially bound by HIF-2. Conversely, TCF12 and JUND are two transcription factors that overlap HIF-2, but not HIF-1, binding sites. Again heatmap analysis corroborate this finding, demonstrating greater TCF12 and JUND ChIP-seq signal within regions preferentially bound by HIF-2 than in regions preferentially bound by HIF-1.

BHLHE40 is a transcription factor that overlaps both HIF-1 and HIF-2 preferentially bound sites separately, as well as showing significant overlap with the entire set of HIF-1 β sites, suggesting that it indiscriminately associates with both HIF-1 and HIF-2 binding sites. Consistent with this, heatmap analysis of BHLHE40 ChIP-seq signal showed comparable signal in both HIF-1 and HIF-2 regions.

Although not exhaustive, these data provide an indication of the spectrum of transcription factors that preferentially associate with HIF-1 and HIF-2 binding sites in HepG2 cells, in addition to factors that indiscriminately associate with both HIF-1 and HIF-2 binding sites. Importantly, after assessing the enrichment of transcription factor motifs within HIF-1 and HIF-2 binding sites, it is clear that a number of these motif enrichments are reflected by actual transcription factor binding site overlap. A prime example of this is the AP-1 transcription factor, whereby the AP-1 motif was shown to be preferentially

enriched within HIF-2 binding sites relative to HIF-1, and moreover AP-1 components such as JUND, FOSL2 and JUN all significantly overlapped HIF-2, but not HIF-1, binding sites. This, in conjunction with prior studies detailing functional associations between HIF and AP-1 (Laderoute, 2005; Villar et al., 2012), suggested the possibility of a genome-wide cooperative relationship between HIF-2 and AP-1, which I will return to in Chapter 5 of this thesis.

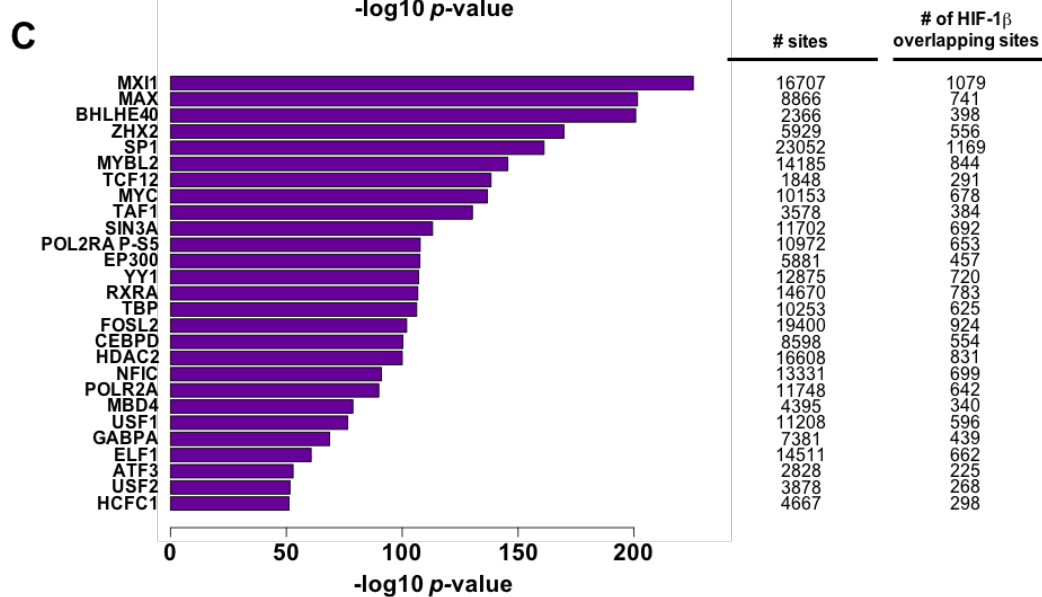
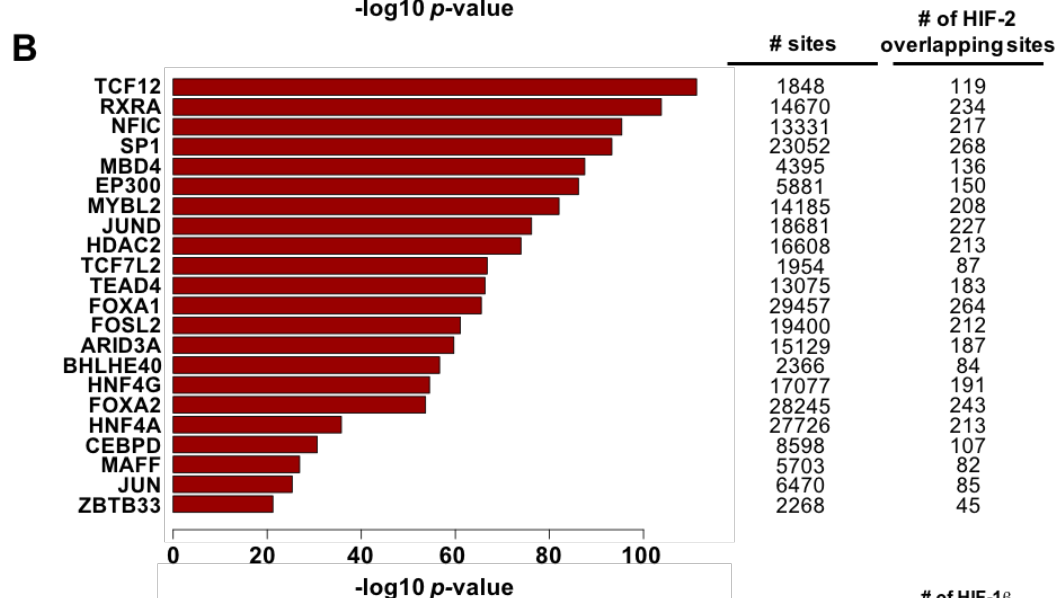
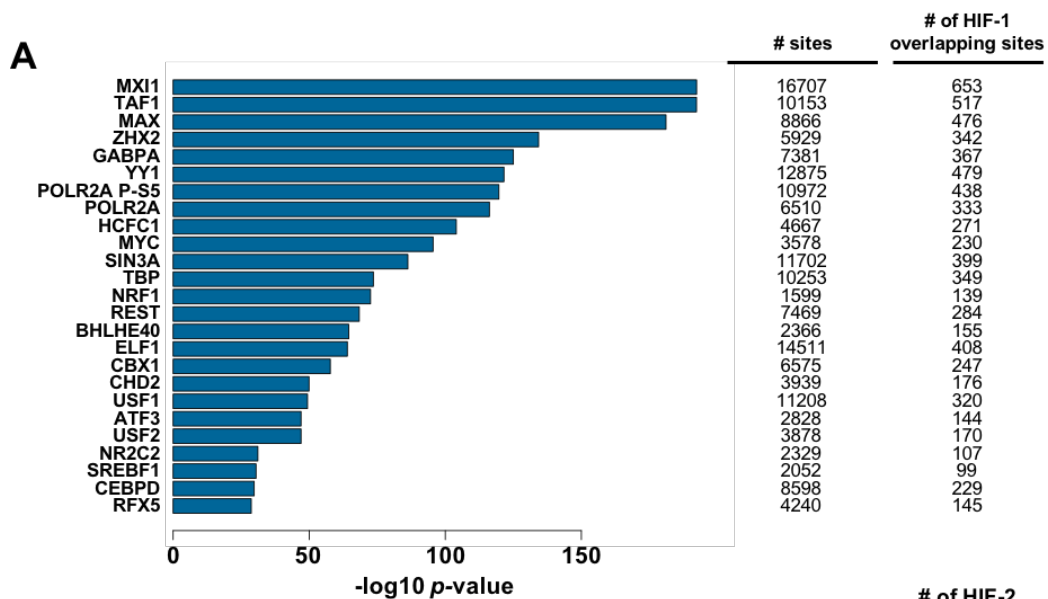


Figure 10. Overlap of HIF binding sites in HepG2 cells with sites identified within ENCODE transcription factor ChIP-seq experiments.

(A) Normoxic ChIP-seq peaks from all transcription factor ChIP-seq experiments performed by ENCODE in HepG2 cells were pooled then overlapped with HIF-1 preferentially bound sites in the same cell line. Bar chart represents the transcription factors that overlap HIF-1 preferentially bound sites to greatest degree of statistical significance, as according to a hypergeometric test. Transcription factors whose sites overlap HIF-1 preferentially bound sites with a $-\log_{10} p$ -value of >30 are shown. The number of peaks called in each ENCODE transcription factor ChIP-seq experiment are shown. Likewise, the number of sites within those peaks that overlap HIF-1 preferentially bound peaks are also shown.

(B) Bar chart as in (A), however sites preferentially bound by HIF-2 have been overlapped with the pooled set of ENCODE transcription factors.

(C) Bar chart as in (A), however the totality of sites bound by HIF-1 β in HepG2 cells have been overlapped with the pooled set of ENCODE transcription factors.

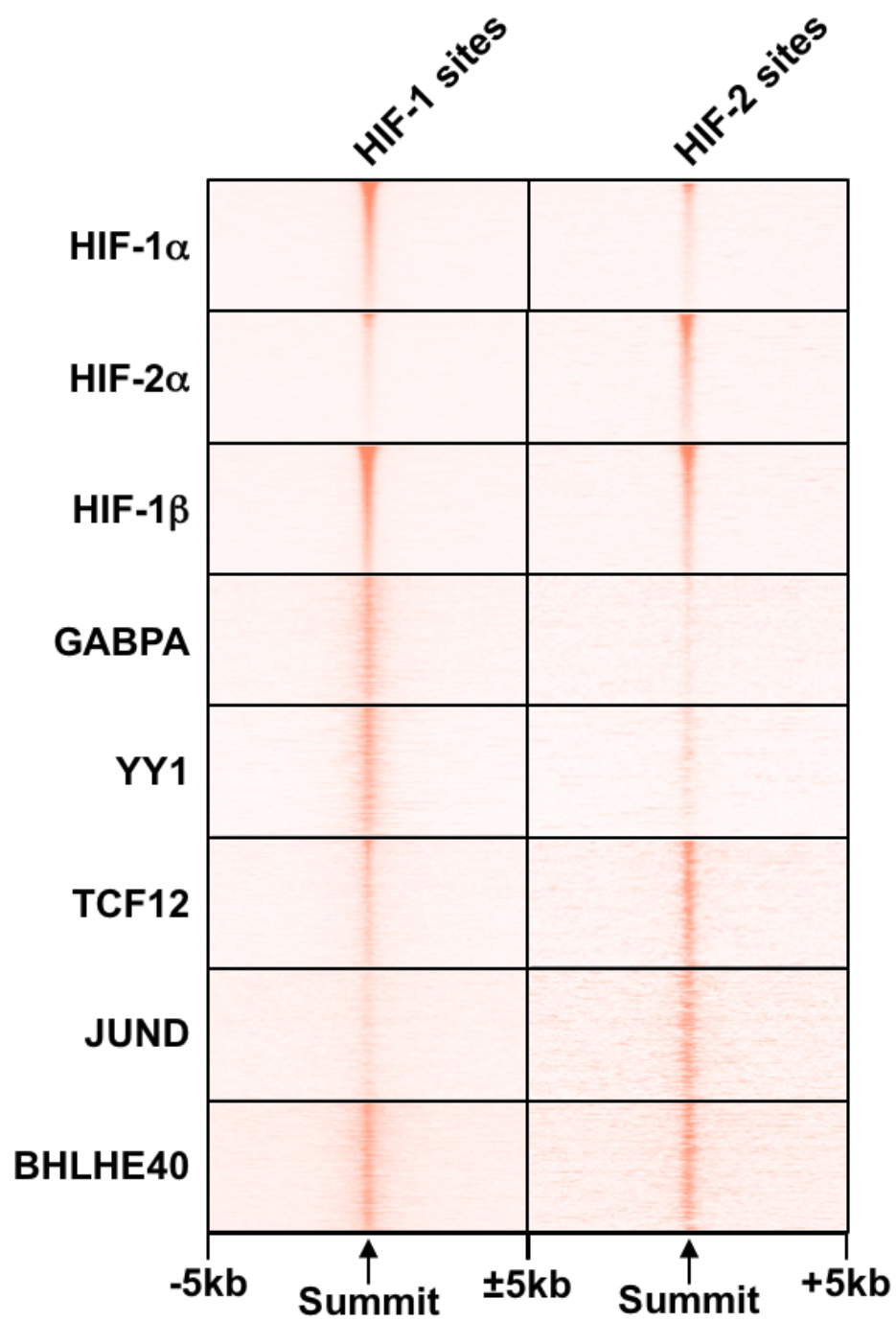


Figure 11. Corroboration of ENCODE transcription factor ChIP-seq peak overlap data via assessment of ChIP-seq signal intensity.

Heatmaps showing HIF-1 α , HIF-2 α , HIF-1 β , GABPA, YY1, TCF12, JUND and BHLHE40 ChIP-seq signal (CPM as colour intensity) for regions $\pm$ 5kb around HIF-1 (1748) and HIF-2 (476) peak summits in HepG2 cells (x-axis). Peaks are sorted by either HIF-1 α or HIF-2 α RPKM (vertical axis) in HepG2 cells. The height of each heatmap vertical axis has been scaled so that the height of HIF-1 and HIF-2 heatmaps are equivalent, in spite of their being a different number of peak summits (vertical axis) in each heatmap. For HIF ChIP-seq experiments, HepG2 cells were incubated for 16 hours in 0.5% O₂ prior to harvesting. All ENCODE ChIP-seq experiments were performed in normoxia. HIF ChIP-seq experiments were performed in biological duplicate, and peaks were only considered if called in both independent biological replicates. Merged .bam files were obtained by combining two biological replicates. These merged .bam files were used to map ChIP-seq signal intensity (CPM) across the defined regions depicted by the heatmaps. For ease of comparing signal intensity in HIF-1 vs. HIF-2 preferentially bound sites, when mapping the reads of an individual ChIP-seq experiment against regions preferentially bound by HIF-1 or HIF-2, the colour scale was maintained equivalent in both heatmaps. The intensity of ChIP-seq signal is not comparable within respective HIF-1 and HIF-2 sets of heatmaps due to differing colour scales being implemented.

3.2.4 Association of sites preferentially bound by HIF-1 or HIF-2 with ENCODE histone modification ChIP-seq data

Having identified which transcription factors bind at HIF-1 and HIF-2 preferentially bound sites, I next wished to assess whether these same HIF-1 and HIF-2 sites associated with specific histone marks. Whilst interacting transcription factors may be involved in the targeting of HIF-1 and HIF-2 to specific binding sites, the histone environment may also be playing a role in HIF binding site selection, as has been reported for other transcription factors (Farnham, 2009).

Raw sequencing data from the ENCODE Consortium ChIP-seq analysis of histone modification in normoxic HepG2 cells was downloaded and processed using the same ChIP-seq analysis pipeline used above (discussed in Methods 7.7.1) Accession numbers for ENCODE data are listed in Appendix 8.1. Genomic regions in HepG2 cells preferentially bound by either HIF-1 or HIF-2 were sorted according to HIF-1 α and HIF-2 α ChIP-seq signal (RPKM) and heatmaps were generated for histone modification ChIP-seq signal as before.

ChIP-seq signal for most histone marks was depleted in the regions that coincided with HIF ChIP-seq peak summits. However, regions immediately flanking the HIF ChIP-seq peak summits showed varying patterns of enrichment for different histone modifications at HIF-1 and HIF-2 preferentially bound sites. H3K27ac, a mark of active enhancers (Creyghton et al., 2010), exhibited greater ChIP-seq signal (CPM) flanking HIF-2 peak summits relative when compared to HIF-1 peak summits. H3K27me3, a mark of inactive enhancers and promoters (Creyghton et al., 2010), was observed to be more prominently flanking HIF-1 summits relative to HIF-2, although the total signal observed is relatively weak. H3K9ac enrichment is associated with low nucleosomal density in the vicinity of

the TSS of a gene (Nishida et al., 2006), and has been shown to be important for the recruitment of TFIID (Transcription factor II D) (Agalioti et al., 2002). A greater enrichment of H3K9ac signal around the HIF-1 α summits is evident relative to the HIF-2 summits. H3K4me1, a mark of enhancers (Heintzman et al., 2007), is enriched to a greater degree around HIF-2 summits relative to that of HIF-1. Whereas conversely, H3K4me3, a mark of promoters (Heintzman et al., 2007), is far more strongly enriched around HIF-1 summits than HIF-2. H3K4me2 is typically associated with active and accessible regions of chromatin (Heintzman et al., 2007), and appears to enrich similarly at HIF-1 and HIF-2 preferentially bound sites. H3K79me2 and H3K36me3 are both associated with gene bodies, however H3K79me2 is more typically associated with the 5' end, and H3K36me3 with the 3' end (Li et al., 2007). H3K79me2 was more strongly enriched at HIF-1 summits, whereas H3K36me3 exhibited an equally weak signal at HIF-1 and HIF-2 sites.

Taken together, there are distinct associations between HIF-1 and HIF-2 preferentially bound sites, and different histone marks, illustrating potential for the histone environment to influence HIF binding site selection.

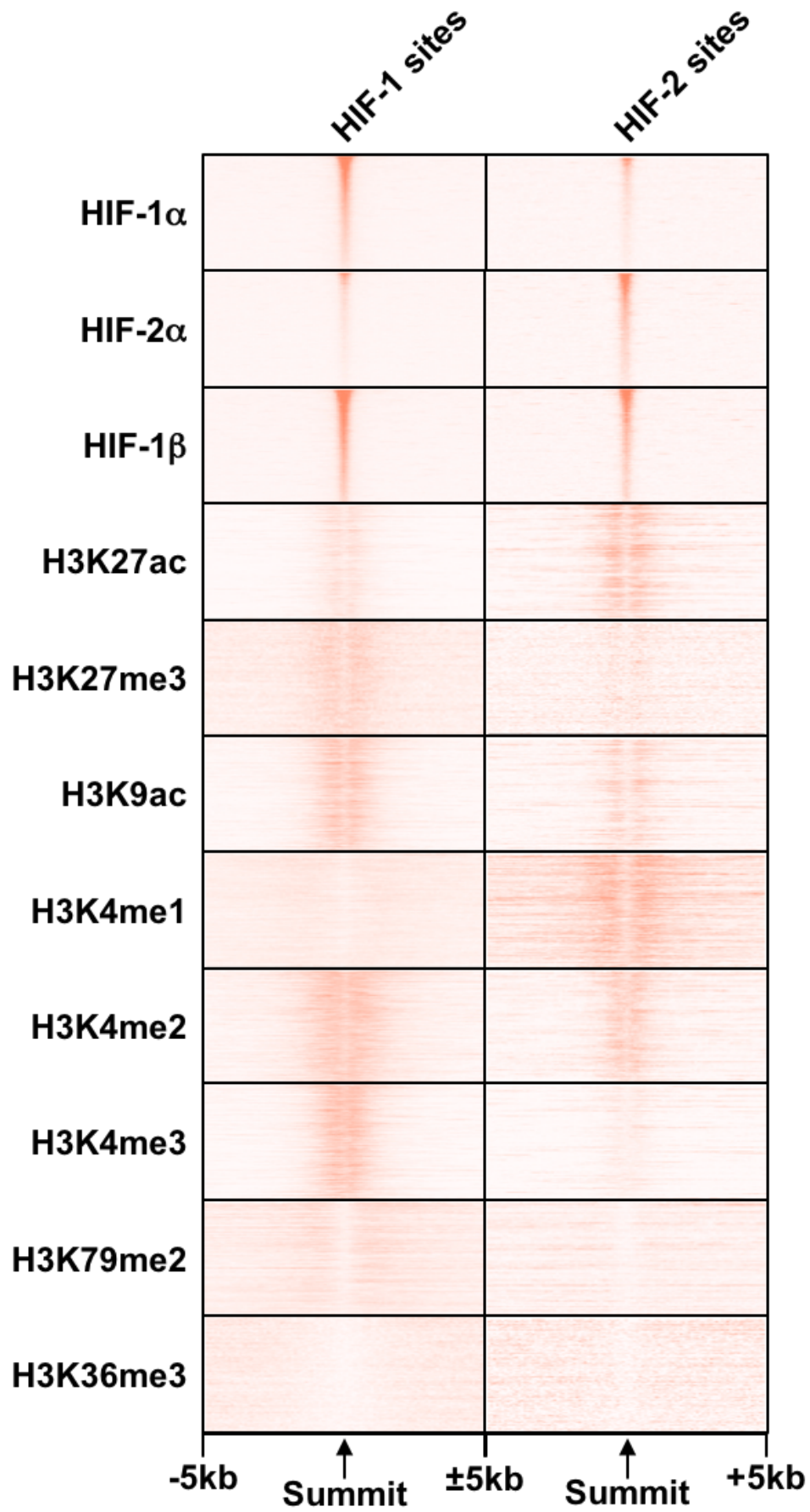


Figure 12. Enrichment of histone marks around HIF binding sites.

Heatmaps showing HIF-1 α , HIF-2 α , HIF-1 β , H3K27ac, H3K27me3, H3K9ac, H3K4me1, H3K4me2, H3K4me3, H3K79me2 and H3K36me3 ChIP-seq signal (CPM as colour intensity) for regions ± 5 kb around HIF-1 (1748) and HIF-2 (476) peak summits in HepG2 cells (x-axis). Peaks are sorted by either HIF-1 α or HIF-2 α peak height (number of overlapping reads at the point of maximum signal within the peak) (vertical axis) in HepG2 cells. The height of each heatmap vertical axis has been scaled so that the height of HIF-1 and HIF-2 heatmaps are equivalent, in spite of their being a different number of peak summits (vertical axis) in each heatmap. For HIF ChIP-seq experiments, HepG2 cells were incubated for 16 hours in 0.5% O₂ prior to harvesting. All ENCODE ChIP-seq experiments were performed in normoxia. HIF ChIP-seq experiments were performed in biological duplicate, and peaks were only considered if called in both independent biological replicates. Merged .bam files were obtained by combining two biological replicates. These merged .bam files were used to map ChIP-seq signal intensity (RPM) across the defined regions depicted by the heatmaps. For ease of comparing signal intensity in HIF-1 vs. HIF-2 preferentially bound sites, when mapping the reads of an individual ChIP-seq experiment against regions preferentially bound by HIF-1 or HIF-2, the colour scale was maintained equivalent in both heatmaps. The intensity of ChIP-seq signal is not comparable within respective HIF-1 and HIF-2 sets of heatmaps due to differing colour scales being implemented.

3.3 Discussion

3.3.1 Cell-type specificity

A previous systematic review of pan-genomic microarray analyses of hypoxia-regulated genes has revealed a high degree of cell-type specificity with relatively few genes regulated across all cell-types examined (Ortiz-Barahona et al., 2010). Specifically, approximately half (11,673) of all protein-coding genes were regulated by hypoxia in one or more cell lines, whilst only 2.22% (259) were regulated by hypoxia in all 14-cell lines. Whilst much of this cell-type specificity may be due to differences in the HIF pathway and HIF isoform expression, additional cell-specific hypoxia-regulated pathways might also influence hypoxic gene regulation in some cell types but not others and contribute to this heterogeneity. Nevertheless, ChIP-seq analysis of pan-genomic HIF binding across 4 cell lines confirmed similar, extensive cell-type differences in binding sites. Although, as described above, the nature of this type of analysis, whereby sites are defined as binding or not in a binary fashion, tends to underestimate the degree of overlap. This data provides the first systematic comparison of ChIP-seq derived HIF-binding sites across multiple cell lines. Furthermore, despite these caveats regarding binary binding site definition, it still allows some important conclusions to be drawn.

Firstly, conserved sites tended to lie closer to gene promoters, while cell-type specific sites more frequently lie further from documented transcriptional start sites. Whilst the current work is the first study of HIF-binding site conservation between cell-types, other studies have examined cell-specific and ubiquitous enhancers more generally and shown similar differences in distance from the nearest TSS (Long et al., 2016). Overall, the more distant distribution of

cell-specific HIF-binding sites from the TSS makes it less likely that these enhancers will be discovered simply by analysis of the promoters and emphasises the utility of pan-genomic techniques in discovering such sites.

Secondly, HIF-2 binding sites exhibit a greater degree of cell-type specificity than do HIF-1 sites, which are more often shared between cell types. This may in part reflect differences in the binding distribution of each isoform around the TSS (HIF-1 binds more frequently closer to the promoter). However, these associations do not infer causality and it may be that HIF-1 binds closer to the promoter because it binds more ubiquitous sites. It is also interesting to note that these differences in cell-specificity of HIF-1 and HIF-2 binding sites also reflect the expression of each isoform in tissues, with HIF-2 expression being restricted to a limited number of cell-types whilst HIF-1 expression is almost ubiquitous. Both of these findings indicate a more general role for HIF-1 in the cellular response to hypoxia, with HIF-2 directing more specialised responses.

3.3.2 Pan-genomic distributions of sites preferentially bound by HIF-1 or HIF-2

Despite their sequence homology and common core consensus HRE motif (RCGTG) (Wenger et al., 2005), HIF-1 and HIF-2 bind distinct, only partially overlapping sets of sites on chromatin (Mole et al., 2009; Schodel et al., 2011). Specifically, HIF-1 preferentially binds at sites closer to the promoter, while HIF-2 binds more frequently at promoter-distant sites. Importantly, these isoform-specific differences in distribution around the local promoter are observed in multiple cell types. This suggests that the observed differences in binding distribution are inherent properties of each isoform and I will go on to test this

hypothesis in the next chapter. In addition, these isoform-specific distributions are consistent between cell types, despite considerable heterogeneity in the binding sites themselves. This implies that the factors governing isoform-specific binding and cell-type-specific binding are not the same and indicates a previously unrecognised level of complexity in the factors governing binding of HIF to specific HRE motifs.

It is well recognised that specific transcription factors may occupy either promoter-proximal or promoter-distal binding distributions. However, it is less clear whether this has been described for closely related transcription factors within the same family. It would therefore be of interest to perform a systematic analysis of binding site distribution of related families of transcription factors using ENCODE and other publically available datasets to determine the specificity or generalisability of this phenomenon.

Neither the distribution of HIF-1 nor the distribution of HIF-2 is comparable to that of accessible HRE motifs suggesting the presence of either positive or negative factors that influence the binding distribution of both isoforms. Given that HIF-1 and HIF-2 share a common HIF-1 β subunit, it is likely that these differences in binding distribution result from specific sequences within the HIF- α subunits. Furthermore, studies in 786O cells indicate that the differences in binding distribution are independent of levels of overexpression. It is therefore possible that this property might be amenable to dissection using a series of domain swapping constructs to express chimeric HIF- α proteins in order to observe the effect of specific HIF- α domains on binding site distribution.

The relative distributions of HIF-1 and HIF-2 binding sites is a remarkable property of HIF-1 and HIF-2. Particularly considering that both HIF-1 and HIF-2

binding sites differ from distribution of the totality of accessible HREs, thus indicating that HIF-1 and HIF-2 are not choosing their binding sites at random. The fixed nature of these binding site distributions across different cell types suggests that there must be distinct sets of factors common to a wide number of – or even all – mammalian cell types that are directing HIF-1 α and HIF-2 α binding site selection. Potential factors influencing these trends will be discussed in subsequent sections of this Discussion.

3.3.3 Outside of the HRE, distinct sets of transcription factor binding motifs enrich within HIF-1 and HIF-2 sites, respectively

HIF-1 α and HIF-2 α exhibit their greatest degree of sequence divergence in their C-terminal domains away from their DNA-binding domain (Ema et al., 1997). Furthermore, detailed analysis of the core consensus motif binding each isoform and of sequences immediately flanking these motifs has not identified significant differences (Wenger et al., 2005; Schodel et al., 2011). Taken together, this suggests that specific interactions between each HIF isoform and additional chromatin-binding proteins that bind beyond the HRE motif may play a role in directing isoform-specific patterns of HIF-binding. Indeed, this hypothesis is supported by my observations.

Firstly, my data has corroborated the published findings (Wenger et al., 2005; Schodel et al., 2011) that HIF-1 and HIF-2 both bind the same consensus motif. Regardless of the cell type being assayed, within both HIF-1 and HIF-2 preferentially bound sites, the HRE is the most significantly enriched motif.

Secondly, I found distinct differences in the enrichment of additional transcription factor motifs at sites preferentially bound by each isoform. This

analysis was facilitated by focusing only on preferentially bound sites and not using all HIF-1 or HIF-2 bound sites, which would include many sites that bound both isoforms equally.

Prior studies have elegantly demonstrated the power of motif analysis in order to identify transcription factor associations on a pan-genomic scale (Biddie et al., 2011; Zaret and Carroll, 2011; Ji et al., 2012; Grontved et al., 2013; Zanconato et al., 2015). A classic example is that of the groundbreaking discovery that the forkhead box (FOX) transcription factor motif was identified in estrogen receptor binding sites, and that the FOXA1 transcription factor was required for estrogen receptor to bind its DNA consensus sequence within chromatin (Carroll et al., 2005). Uncovering distinct families of transcription factor motifs within regions preferentially bound by either HIF-1 or HIF-2 will enable me to argue that these families of transcription factors are contributing to the relative binding site selectivity of HIF-1 and HIF-2, thereby providing justification to experimentally assess these postulations.

A consistent pattern within HIF-1 α preferentially bound sites across all cell lines assayed is the enrichment of GC-rich motifs. Given that CpG islands tend to reside within promoter regions (Saxonov et al., 2006), and given that I have previously outlined HIF-1 α 's preferential association with promoter regions, it is not surprising that GC-rich motifs are enriched within HIF-1 α binding sites. However, a key question is whether the GC-rich motifs, in conjunction with the transcription factors that bind these motifs, are what directs HIF-1 α to these loci. Or, whether the enrichment of these motifs is a secondary occurrence as a result of HIF-1 α choosing to bind in the local histone environment.

Many of the transcription factors whose motifs are enriched within HIF-1 α sites have been functionally associated with HIF-1 α previously. For instance, perturbation of SP-1, a zinc-finger transcription factor, binding to sites neighbouring HIF-1 α binding sites has been shown to augment the capacity of HIF-1 α to induce expression of *MDR1*, *RORA*, *CD147*, *CA9* and *GAPDH* (Koizume and Miyagi, 2015). Another zinc-finger transcription factor, KLF5, has been functionally linked to HIF-1 α via its crucial role in hypoxia-induced vascular remodeling (Li et al., 2016). ZNF263 has not previously been functionally associated with HIF-1 α . The palindromic E-box motif may be bound by a huge array of bHLH domain containing transcription factors (Jones, 2004), including MYC, MAX, MXI1, BHLHE40, BMAL and CLOCK. Notably, an E-box motif (CACGTG) may contain a HRE, therefore in some circumstances, HIF-1 could feasibly be in competition with E-box binding proteins for binding sites. Nevertheless, in a number of instances, HIF-1 α has been reported to act in concert with MYC to induce the expression of *NBN*, *CDKN1A* and a wide array of glycolytic genes (Gordan et al., 2007; Dang et al., 2008).

Sites preferentially bound by HIF-2 α clearly enrich for a distinct set of motifs relative to HIF-1 α . Specifically in HepG2 cells, HIF-2 α appears to enrich strongly for the forkhead box (FOX) motif. The FOX proteins are formed of 50 family members that can each affect a wide array of cellular behaviours via influencing the expression of an extremely broad spectrum of genes (Benayoun et al., 2011; Lam et al., 2013). No prior studies have demonstrated a direct interaction between HIF-2 α and FOX proteins, although it has been demonstrated that FOXO3 may induce the expression of *CITED*, which can subsequently repress the ability of HIF-1 α to coordinate the transcriptional

activator, p300 (Bakker et al., 2007). In HKC8 and RCC4 cells, sites preferentially bound by HIF-2 α displayed a remarkably prominent enrichment for the 12-O-Tetradecanoylphorbol-13-acetate (TPA) response element (TRE). This motif is bound by a dimer comprising components within two AP-1 families. Specifically, JUN family members (c-JUN, JUND, JUNB) may homodimerise, or heterodimerise with FOS family members (c-FOS, FOSB, FOSL1, FOSL2) in order to engage the TRE (Lee et al., 1987; Eferl and Wagner, 2003). There have been numerous reports functionally associating AP-1 and HIF (Laderoute, 2005; Villar et al., 2012). Moreover, the Ratcliffe Group previously reported that HIF-1 α and HIF-2 α canonical binding sites in MCF7 both enriched for the TRE (Schodel et al., 2011), and overexpressed HIF-2 α binding sites in 7860 cells exhibited a highly significant enrichment for the TRE (Salama et al., 2015). It should be noted that the TRE did enrich within HIF-2 α binding sites in HepG2 cells, although not nearly to the same significance as FOXK2. Conversely, the FOX motif enriched within HIF-2 α sites in HKC8 cells, but again, not to the same significance as the AP-1 motif.

Taken together, these results illustrate distinct transcription factor motifs that preferentially enrich within HIF-1 α and HIF-2 α binding sites. However, for any transcription factor whose motif I have shown to be enriched with sites preferentially bound by either HIF-1 α or HIF-2 α , there has been no pan-genomic intervention-based study illustrating widespread influence on HIF binding capabilities or binding site selection. Whilst there are published studies have utilised motif analysis to identify transcription factors that may be cooperating with their transcription factor of interest, there are caveats that should always be considered when interpreting motif enrichment. As discussed earlier, certain

motifs are inherently more likely to be enriched within certain genomic regions owing to the bias toward particular nucleotides. GC-rich motifs are more likely to be present in promoter regions given the preponderance of CpG islands in promoters (Saxonov et al., 2006). Furthermore, the presence of motif does not necessarily translate to transcription factor binding. And finally, multiple transcription factors within a given transcription family may be capable of binding the same transcription factor motif. Therefore, formulating a transcription factor cooperativity- or competition-based hypothesis on the basis of motif enrichment alone leads to an increased likelihood of an irrelevant transcription factor being chosen. For that reason, a more robust method for identifying co-operating transcription factors is to couple motif analysis to pan-genomic assays of transcription factor binding site occupancy such as ChIP-seq.

3.3.4 Distinct sets of transcription factor binding sites coincide with HIF-1 and HIF-2 binding sites

Analysis of the overlap between transcription factor ChIP-seq datasets generated by the ENCODE Project in normoxic HepG2 cells, with HIF-1 and HIF-2 preferentially bound sites, largely corroborated the enrichment of transcription factor binding motifs within the preferentially bound HIF sites. For example, the E-box motif was enriched within HIF-1 binding sites and E-box binding transcription factors are among the factors most enriched at these sites. Specifically, MXI1, MAX, MYC and BHLHE40 are the 1st, 3rd, 10th and 15th most significantly overlapping factors, respectively. Interestingly, MXI1 and BHLHE40 have both been described as transcriptional repressors (Adhikary and Eilers, 2005; Kato et al., 2014). This suggests that repressive bHLH transcription factors

may be binding and repressing HIF target loci prior to its stabilisation in hypoxia. Binding of HIF might compete with and displace binding of these repressive factors (since E-boxes are also HREs) leading to an enhanced dynamic range in gene expression between normoxia and hypoxia.

Interestingly, both *MXI1* and *BHLHE40* themselves are also HIF-target genes (Giatromanolaki et al., 2003; Corn et al., 2005) and this may have important implications for the temporal response to hypoxia. Specifically, if HIF-1 does out-compete and displace MXI1 and BHLHE40 from target loci during the onset of hypoxia, the subsequent induction of *MXI1* and *BHLHE40* expression may reverse this process and serve as a negative feedback mechanism.

In addition to the bHLH proteins, MXI1 and BHLHE40, a number of other HIF target genes act as repressive transcription factors. For example ZHX2 and SIN3A (Kawata et al., 2003; Grzenda et al., 2009), were both hypoxia inducible in my HepG2 RNA-seq analysis. Their binding sites also overlap (at least partially) with HIF-1 binding sites. Thus it is possible that temporal suppression of HIF target genes as well as hypoxic repression of non-HIF-binding genes is a general feature of hypoxia-inducible transcriptional repressors. The overlap with HIF-1 preferentially bound sites rather than HIF-2 bound sites might also reflect important temporal differences in the function of the two isoforms. Induction of HIF-1 α in hypoxia is generally held to be more transient than for HIF-2 α and this coupled with induction and subsequent binding of transcriptional repressors at HIF-1, but not HIF-2 sites, would be consistent with a more transient role for HIF-1 in the hypoxic response, whilst the HIF-2 response is maintained for longer (Tian et al., 2011).

The FOX motifs were the most striking motifs enriched within sites preferentially bound by HIF-2 in HepG2 cells and occupied FOXA1 binding sites also significantly overlapped these sites in the ENCODE ChIP-seq datasets. FOXA1 is pioneer transcription factor that has previously been shown be necessary for the binding of the estrogen receptor to bind its target consensus sequences within chromatin (Zaret and Carroll, 2011). Since the vast majority of HIF binding sites are accessible in normoxia (Schodel et al., 2011), before HIF binds, it is not surprising that pioneer or bookmarking factors are present at HIF-binding sites to preserve this open chromatin state ready for rapid binding of HIF in hypoxia, although promoter-proximal sites might be maintained in other ways.

Among the transcription factors most significantly overlapping HIF-2 binding sites are TCF12 (Transcription factor 12), RXRA (Retinoid X receptor α), NFIC (Nuclear factor I C), SP1 and MBD4 (Methyl-CpG binding domain protein 4). Aside from SP1, none of these factors has previously been functionally associated with HIF.

TCF12 is a bHLH transcription factor that binds an E-box (Zhang et al., 1992). Again, as previously discussed, TCF12 could feasibly be competing with HIF-2 α for HREs.

Interestingly the concordance between SP-1 binding and HIF-2, but not HIF-1 binding is in contradistinction to the relative enrichment of the SP-1 consensus motif, which is more strongly enriched at HIF-1 preferentially bound sites. This again emphasises the difficulties interpreting motif enrichment as opposed to binding site occupancy when looking for differences between HIF-1 and HIF-2 preferentially bound sites. SP-1 binding at HIF-1 binding sites has previously been shown to be necessary for hypoxic-transactivation of some HIF-1

target genes (Koizume and Miyagi, 2015). Whether it does so by facilitating HIF binding and whether effects are amplified at HIF-2 sites compared to HIF-1 sites remains to be determined.

The AP-1 components, JUND, FOSL2 and JUN, in addition to the AP-1 interacting transcription factor, TEAD4 (Zanconato et al., 2015), are all significantly overlapping HIF-2 binding sites in HepG2 cells, whilst not significantly overlapping HIF-1 sites. I observed extremely strong enrichment of the TRE within HIF-2 in RCC4 and HKC8 cells, but only a modest enrichment in HepG2 cells. Nevertheless, such a significant cohort of AP-1 transcription factor overlap suggests there is a strong possibility of a functional significance in the overlapping transcription factor binding sites. AP-1 has been shown to potentiate chromatin accessibility, in order to enable DNA binding of the glucocorticoid receptor and TEAD (Biddie et al., 2011; Zanconato et al., 2015), thus enabling transactivation of their respective sets of target genes. AP-1 may be performing an analogous role with respect to HIF-2 and I will return to this in a later section of my thesis.

The transcriptional co-activator and histone acetyltransferase p300/CBP was significantly enriched at HIF-2 preferentially bound sites and at common sites, but less so at HIF-1 specific sites. This might be expected since p300/CBP is known to physically associate with the C-TAD of HIF-1 α and HIF-2 α and this interaction is inhibited by oxygen-dependent hydroxylation of the C-TAD by FIH. However, the ENCODE datasets upon which this analysis is based were performed in normoxia. Thus p300/CBP is present at HIF-binding sites even before HIF binds and this may account for the presence of H3K27ac at these sites in normoxia. Since p300/CBP is recruited by many different transcription

factors, it would suggest the presence of additional transcription factors at or close to HIF-binding sites in normoxia. Indeed, consistent with this hypothesis, most HIF target genes are already expressed to a degree in normoxia. Given the physical association between HIF and p300/CBP together with the increase in H3K27ac observed at HIF-binding sites upon HIF binding it is likely that the p300/CBP binding signal would also increase in hypoxia at these sites.

Similarly, HDAC2 (Histone deacetylase 2) was also observed to be significantly overlapping HIF binding sites. It is not known what happens to HDAC2 binding in hypoxia, but it seems likely that the resultant H3K27ac signal is a result of the dynamic interplay between p300/CBP and HDAC2.

Since there are more HIF-1 binding sites than HIF-2 binding sites in HepG2 cells, it is not surprising that transcription factors that most significantly overlap the totality of HIF-1 β binding sites are somewhat dominated by those factors that overlap HIF-1 binding sites. However, of note is the significant overlap of BHLHE40. It is overlapping the totality of HIF-1 β sites with far greater significance than it significant overlaps HIF-1 and HIF-2 sites alone. This suggests that it is not selectively associating with HIF-1 or HIF-2 sites, rather it is indiscriminately associating with both HIF-1 and HIF-2 binding sites.

Whilst all of these associations between transcription factors and HIF binding sites are of interest and provide exciting insights, a major caveat is that all of the ENCODE ChIP-seq data was acquired in normoxia. As indicated above, it is possible that hypoxia and/or HIF-binding re-distributes one or more of these associated factors. Thus it would be interesting to examine the effect of hypoxia on the binding of these factors by performing further ChIP-seq experiments in hypoxia.

3.3.5 Contrasting histone environments at HIF-1 and HIF-2 sites, respectively

HIF-1 was seen to bind chromatin more often in association with histone H3K4me3 modification (a mark of active promoters), whilst HIF-2 was associated with the H3K4me1 chromatin mark (an indicator of active enhancers) (Heintzman et al., 2007). Both isoforms bound in association with marks of active chromatin (Creyghton et al., 2010) such as H3K27ac (discussed above) and were depleted of repressive marks. Given the propensity for HIF-1 to bind close to the promoter and for HIF-2 to bind to promoter-distant enhancers, the observed differences in associations with histone marks are not surprising. However, what is in question is the extent to which these differences simply reflect the differences in binding distribution or to what extent they are responsible for them. In other words, transcription factors may recognise their respective binding motifs only in the context of specific histone modifications. Interventional studies to dissect these possibilities would be challenging to interpret, since they would have widespread pan-genomic effects on histone modification, which would alter the fundamental identity of the cell leading to many indirect effects. Recent advances in repurposing the CRISPR-Cas9 system to target histone modifying activity, rather than nuclease activity, to specific genomic locations might provide useful tools in this respect but remain an immature technology.

It is also possible that HIF-1 and HIF-2 interact indirectly with different histone modifications through intermediate chromatin “readers”. These are proteins or complexes containing specific domains that are known to interact with particular histone modifications (Musselman et al., 2012). Relatively few readers of H3K4me1 are known. Those that are known comprise the chromo-barrel

domain containing proteins and double chromodomain containing (DCD) proteins. Interestingly, DCDs recognise H3K4me1, me2 and me3 therefore if a histone modification reader is directing HIF-2 to H3K4me1-rich regions it is more likely to be a chromo-barrel domain-containing protein such as TIP60 rather than a DCD protein. TIP60 has not previously been associated with HIF-2. However, somewhat paradoxically, a recent study illustrated that TIP60 may function as a co-activator of HIF-1 (Perez-Perri et al., 2016), although it was HIF-1 α that directed TIP60 binding rather than vice versa.

Plant homeodomain (PHD), tandem Tudor domain (TTD) and DCD containing proteins, in addition to the CW-type zinc finger, are all readers of H3K4me3 (Musselman et al., 2012). Again, DCDs are unlikely to be responsible for specifically recruiting HIF-1 to H3K4me3-annotated regions, since DCDs also bind H3K4me1. TAF3 contains a PHD domain and is a subunit of the TFIID complex (Vermeulen et al., 2007). I observed a significant binding site overlap between HIF-1 and TAF1, another subunit of the TFIID complex, so a direct or indirect interaction with TAF3 may be partially involved in directing HIF-1 to its preferentially bound sites.

Beyond co-localisation on chromatin, an association between HIF isoforms and reader proteins could be examined using an affinity purification mass-spectrometry (AP-MS) approach. To date this approach, when applied to HA-tagged HIF-2 α , has not identified any proteins containing chromo-barrel domains (Steunou et al., 2013). Whilst these approaches are technically challenging, a number of studies that have used formaldehyde crosslinking and nuclear extraction prior to AP-MS have yielded successful results in the context of the

estrogen receptor and FOXK2 (Mohammed et al., 2013; Ji et al., 2014). A similar methodology could be attempted, in the future, for HIF-1 α and HIF-2 α .

3.3.6 Conclusion

In conclusion, HIF-1 and HIF-2 occupy distinct and only partially overlapping sets of binding sites. HIF-1 preferentially binds sites close to the promoter, whilst HIF-2 binds more frequently to promoter-distant enhancers. HIF-1 sites are more likely to be conserved between different cell-types, whilst HIF-2 binding sites are more often cell-specific. The overall patterns of HIF-binding with respect to the local promoter are preserved between cell lines despite the binding sites themselves differing to a large extent. Beyond the common consensus HRE motif, each isoform is associated with specific accessory transcription factor binding motifs, binding site occupancy and specific histone modifications. The extent to which these and the presence of the other HIF isoform influence binding site selection of HIF-1 and HIF-2 remain unclear and form the basis of the next two chapters of my thesis.

Chapter 4: Insights into the pan-genomic binding site plasticity of HIF-1 and HIF-2, and the functional consequences of shared HIF- α binding sites

4.1 Introduction

Hypoxia inducible factors -1 α and -2 α both heterodimerise with their canonical dimerisation partner, HIF-1 β , in order to bind hypoxia response elements and transactivate target genes in hypoxia (Kaelin and Ratcliffe, 2008) (Henceforth, I shall refer to HIF-1 when discussing a canonical HIF-1 α /1 β heterodimer, and HIF-2 when discussing a canonical HIF-2 α /1 β heterodimer). Early pan-genomic HIF studies have shown that HIF-1 and HIF-2 bind overlapping as well as distinct sets of sites (Mole et al., 2009; Schodel et al., 2011). However, in the previous chapter, I demonstrated that HIF binding sites may be stratified via a more quantitative approach, subsequently leading to the identification of sites that are ‘preferentially’ bound by HIF-1 or HIF-2, rather than defining sites as distinctly bound by HIF-1 or HIF-2. These pan-genomic studies beg the question whether the spectrum of HIF-1 or HIF-2 binding sites, and their respective binding strength at a given locus, are influenced by the presence of the other HIF- α isoform in the cell? Specifically, are HIF-1 and HIF-2 capable of binding new sites to compensate for the absence of a HIF- α isoform, or does the absence of a HIF- α isoform bear no influence on the binding spectrum of the other HIF- α isoform?

To date, the only study that has alluded to the effects that one HIF- α isoform may have on the pan-genomic binding of the other, is one that overexpresses HIF-1 α in 786O cells and performs ChIP-seq on exogenous HIF-1 α and endogenous HIF-2 α (Salama et al., 2015). This study demonstrated that an 8-fold overexpression of HIF-1 α had no effect on the pan-genomic distribution of endogenous HIF-2 α binding. 786O cells harbour a mutated form of *HIF1A* whereby no mRNA that forms the HIF-1 α transactivation domain is detectable

(Warnecke et al., 2004), therefore it is likely that there is no functional endogenous HIF-1 α protein being expressed (Maxwell et al., 1999). Salama et al reports that HIF-2 α adopts a highly similar promoter-distal binding site distribution both in the presence and absence of overexpressed HIF-1 α .

Whilst no study previously demonstrating whether endogenous HIF-1 and HIF-2 are capable of binding new sites to compensate for the absence of a HIF- α isoform, a related subsidiary question is whether any level of binding redistribution is associated with a maintenance or alteration of HIF target gene activation. In the absence of data concerning HIF DNA binding, studies have shown a small capacity of HIF-2 α to compensate for the lack of HIF-1 α in 786O cells by maintaining the expression of a small number of genes that are HIF-1 targets in VHL-competent cells (Shinojima et al., 2007). However, overexpression of a mutant form of HIF-2 α that is stable in normoxia, in the absence of concurrent mutant HIF-1 α overexpression, demonstrated that HIF-2 α could induce a broad spectrum of genes, it was incapable of inducing the typically-HIF-1 regulated glycolytic genes (Hu et al., 2003). These studies suggest that HIF-2 may be capable of exhibiting partial, but not complete, compensation for HIF-1. However, the latter study was performed using an artificial system comprising HIF-2 α stabilised in normoxia, therefore the complete scope of the capacity of HIF-1 and HIF-2 to exhibit compensatory DNA binding and gene transactivation is far from fully elucidated.

In light of the aforementioned studies, I utilised CRISPR/Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats) technology to generate HKC8 (renal proximal tubule) cell lines that are either simultaneously, or separately, deficient in HIF-1 α and HIF-2 α . I then performed HIF-1 α , HIF-2 α and HIF-1 β

ChIP-seq in these cell lines, in addition to RNA-seq in hypoxia and normoxia, in order to understand a number of fundamental questions regarding HIF DNA binding, specifically:

1. Does the absence of HIF-1 α or HIF-2 α affect the pan-genomic distribution of the present HIF- α isoform?
2. Do binding site redistribution events associate with maintenance or alteration of HIF target gene activation?
3. What are the characteristics of a HIF-1 and HIF-2 shared binding site that results in the site regulating a nearby gene in a HIF-1, or HIF-2, dependent manner?

4.2 Results

4.2.1 Validation of CRISPR KO cell line integrity

To determine whether the presence of one HIF- α isoform was able to alter binding of the other, I employed isogenic HKC-8 renal tubule cell lines in which HIF-1 α or HIF-2 α had been ablated (both separately and together) using CRISPR/Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats). These had been designed to target exon 1 of the *HIF1A* gene and exon 2 of the *EPAS1* gene, which code for their DNA binding domains. Non-homologous end joining (NHEJ) resulted in frameshift mutations, that lead to nonsense mediated decay of the *HIF1A* and *EPAS1* transcripts. Single cell clones were all generated from pools of transfected cells by former Ratcliffe Group member, Elizabeth Murray and are referred to as '-1 α KO', '-2 α KO' and 'DKO' respectively.

To establish confidence in the legitimacy of the knockout cells, I performed Western blot analysis on whole cell extracts derived from WT, -1 α KO, -2 α KO and DKO HKC8 cells that had been cultured in normoxia, or in hypoxia (0.5% O₂ for 16 hours) (Figure 1A). HIF-1 α protein was not detectable in hypoxic -1 α KO cells, whereas there was strong induction in both WT and -2 α KO cells. Conversely, HIF-2 α protein was not detectable in hypoxic -2 α KO cells despite strong induction in WT and -1 α KO cells. HIF-1 β protein was constitutively expressed at a constant level across all cell lines in normoxia and hypoxia. Interestingly, there is a modest decrease in HIF-1 α protein in the -2 α KO cells relative to WT, whilst conversely there is a slight increase in HIF-2 α protein in -1 α KO cells relative to WT. This is despite comparable levels of mRNA in the RNA-seq analysis.

I next performed ChIP-qPCR on nuclei from cells cultured in 0.5% O₂ for 16 hours prior to harvesting. Immunoprecipitations were performed using the PM-14 antibody against HIF-1 α , and the PM-9 antibody against HIF-2 α . Enrichment of HIF-1 α and HIF-2 α signal at a non-HIF binding site (*ACTG*) and a HIF-1 α and HIF-2 α common binding site (*NDRG1*) was quantified as a percentage of the input chromatin (Figure 1B). Three biological replicate experiments were performed. In WT cells, both HIF-1 α and HIF-2 α enriched at the *NDRG1* locus to a similar degree. Knockout of each HIF- α isoform abrogated HIF- α ChIP-qPCR signal for that isoform in the -1 α KO, -2 α KO and DKO cells. Interestingly, the ChIP-qPCR signal for the ablated isoform was slightly higher in each of the single knockout cells (7% relative to WT for HIF-1 α and 5% relative to WT for HIF-2 α) than in the double KO cells suggesting that there may be a very small degree of cross reactivity between the HIF-1 α and HIF-2 α ChIP antibodies that had not been detected previously on ChIP-immunoblot (Lau et al., 2007). Whilst this may be significant, it will tend to underplay rather than exaggerate the observed differences in binding of the two isoforms.

In each case, the antibody used for immunoblot and for ChIP was directed against the C-terminus of the HIF- α subunit, so it remains a formal possibility that truncated forms of each molecule that are not recognised by the antibodies are still expressed (as for HIF-1 α in 786O cells). However, the CRISPR target region includes the DNA-binding domain of each isoform, so if they existed, these would be predicted to be non-DNA-binding and thus non-functional from a transcriptional perspective. To establish total confidence in the validity of the CRISPR-defective clones, each HKC8 knockout cell allele containing *HIF1A* and *EPAS1* and must be cloned and sequenced. This work is ongoing.

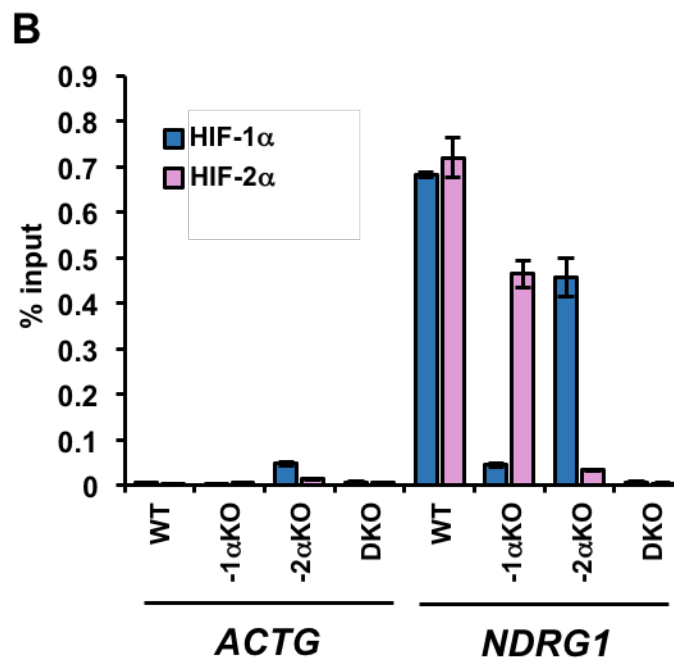
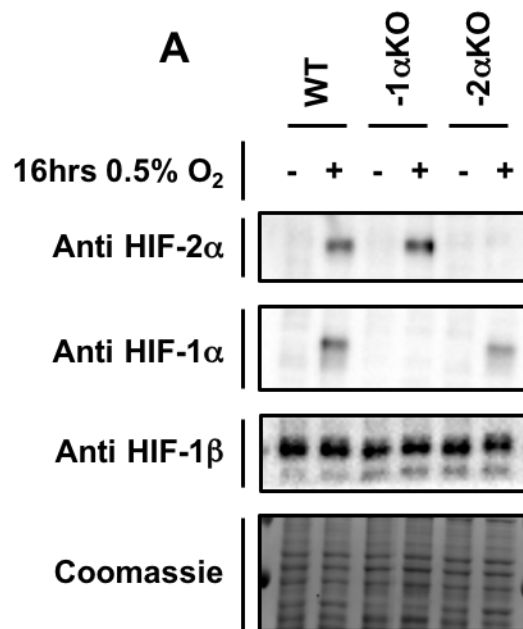


Figure 1. Validation of HKC8 CRISPR HIF knockout cells.

(A) Western blot of HKC8 wild type (WT), HIF-1 α KO and HIF-2 α KO lysates harvest after culturing in normoxic conditions or 16 hours of 0.5% O₂. Coomassie gel illustrating total protein loading.

(B) ChIP-qPCR performed on HIF-1 β , HIF-1 α and HIF-2 α in HKC8 WT, -1 α KO, -2 α KO and DKO cells. Binding enrichment at the *ACTG* and *NDRG1* loci are quantified via the percentage input method. Error bars = $\pm$ Standard deviation for three biological replicates.

4.2.2 HIF binding site distribution and initial assessment of HIF- α binding site plasticity

Having confirmed the effects of the HIF-1 α and HIF-2 α CRISPR interventions on protein expression and DNA binding, I assessed regions bound by either HIF-1 or HIF-2 in WT cells to determine whether any extra sites were subsequently occupied by HIF-2 in -1 α KO cells, or by HIF-1 in -2 α KO cells, respectively. Sites identified by MACS as binding either HIF-1, HIF-2 or both isoforms were first determined using BEDTools intersect (v2.26.0). In WT cells, 872 of the 1,083 HIF-1 binding sites were also occupied by HIF-2 (Figure 2A). However, in the HIF-1 α knockout cells fewer (only 728) rather than more of these sites were occupied by HIF-2 (Figure 2B). 872 of the 3,246 HIF-2 binding sites in WT cells were co-occupied by HIF-1, whilst this number also fell (to 399) in the HIF-2 α KO cells (Figure 2C). These results do not support qualitative re-distribution of one HIF isoform to binding sites of the other when it is knocked out.

To determine whether removal of one HIF- α isoform could lead to the other isoform binding in its place, I next mapped the pan-genomic distributions of HIF-1 and HIF-2 binding sites in the WT, -2 α KO, and -1 α KO cells with respect to the distance between binding sites and the nearest gene. Consistent with the prior observations whereby few sites subsequently became occupied by the remaining HIF- α isoform, no major redistribution of HIF-1 or HIF-2 binding was observed (Figure 2D-E).

Reassuringly, the HIF-1 β binding site distributions in -2 α KO and -1 α KO cells are highly similar to the distributions of the remaining HIF- α (Figure 2F-G). This firstly indicates that the consistently disparate HIF-1 and HIF-2 distributions are not due to the antibodies for ChIP-seq. Moreover, this corroborates the data

from Chapter 3 that in these cells HIF-1 β is only binding DNA in conjunction with HIF- α .

Results described in this section do not support qualitative re-distribution of one HIF isoform to binding sites of the other when it is knocked out. However, it is possible that removal of one HIF-isoform might lead to quantitative changes in binding of the other isoform. Therefore, I next performed quantitative analysis of the changes in HIF-1 α and HIF-2 α ChIP-seq signal at these sites following ablation of each HIF- α isoform.

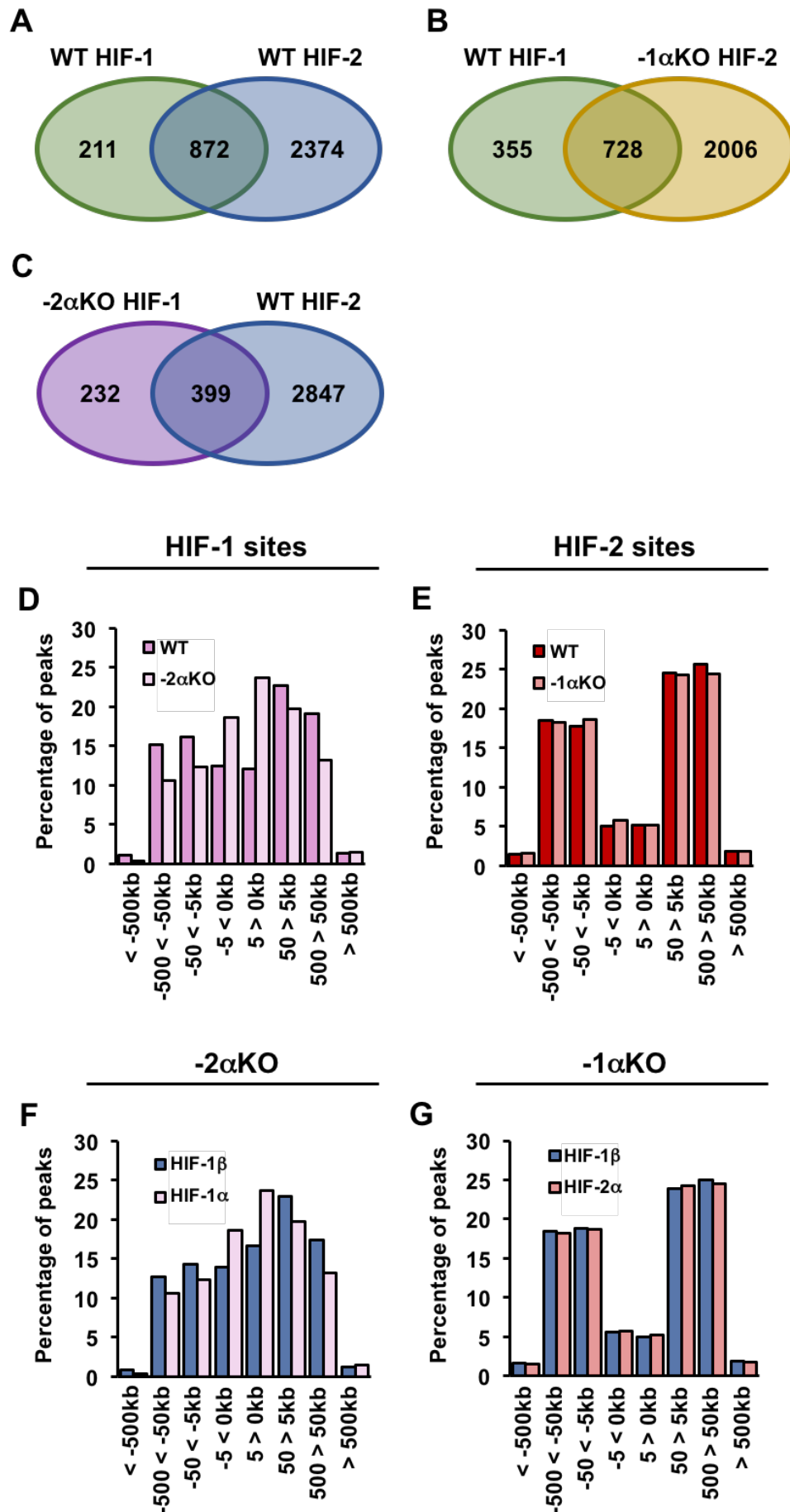


Figure 2. Primary assessment of HIF-1 and HIF-2 re-distribution in HKC8 KO cells relative to WT.

(A) Venn diagram illustrating the number of overlapping peaks between HKC8 WT HIF-1 (green) and WT HIF-2 (blue).

(B) Venn diagram illustrating the number of overlapping peaks between HKC8 WT HIF-1 (green) and HKC8 -1 α KO HIF-2 (gold).

(C) Venn diagram illustrating the number of overlapping peaks between HKC8 WT HIF-1 (blue) and HKC8 -2 α KO HIF-1 (purple).

(D) Histogram depicting the proportion of HIF-1 peak summits in HKC8 WT (pink) and -2 α KO (light pink) cells falling within defined bins based upon the peak summit's distance from the TSS of the nearest gene.

(E) Histogram depicting the proportion of HIF-2 peak summits in HKC8 WT (red) and -1 α KO (light red) cells falling within defined bins based upon the peak summit's distance from the TSS of the nearest gene.

(F) Histogram depicting the proportion of HIF-1 β (blue) and HIF-1 (light pink) peak summits in HKC8 -2 α KO cells falling within defined bins based upon the peak summit's distance from the TSS of the nearest gene.

(G) Histogram depicting the proportion of HIF-1 β (blue) and HIF-2 (light red) peak summits in HKC8 -1 α KO cells falling within defined bins based upon the peak summit's distance from the TSS of the nearest gene.

4.2.3 Differential binding analysis of HIF ChIP-seq in HKC8 HIF- α knockout cells

Three sets of HIF-binding regions were first defined in WT cells as outlined in Figure 2C:

- regions bound by both HIF-1 and HIF-2 in WT HKC8 cells
- regions bound by HIF-1 but not HIF-2 in WT HKC8 cells
- regions bound by HIF-2 but not HIF-1 in WT HKC8 cells

For each region HIF-1 α , HIF-2 α and HIF-1 β ChIP-seq signal (mean RPKM, determined using BEDTools multicov (v2.26.0), across the two biological replicates) was quantified and statistically significant differences were determined using an independent 2-group unpaired t-test. Boxplots were then generated in R to summarise the ChIP-seq signal within each set of regions (Figure 3A-F).

In neither case did HIF-1 α signal increase at HIF-2 sites when HIF-2 α was ablated, nor did HIF-2 α signal increase at HIF-1 sites when HIF-1 α was deleted. This is consistent with the qualitative analysis discussed previously, and again does not support the hypothesis that one isoform can substitute for the other at these sites. Indeed, HIF-1 α signal actually fell upon HIF-2 α ablation ($p < 0.001$; Figure 3E). This was consistent across all 3 sets of sites ($p < 0.001$; Figure 3A,C,E) and may simply reflect the reduction in HIF-1 α protein in these cells.

However, if compensation is possible, then it would most likely be observed at sites that had been identified as capable of binding both isoforms. I therefore examined HIF- α ChIP-seq signals at the set of common binding sites following ablation of each isoform. Again, HIF-1 α signal was reduced rather than increased at these sites in the -2 α KO cells ($p < 0.001$; Figure 3A). Conversely, HIF-2 α did show a statistically significant increase in signal in the -1 α KO cells (p

< 0.05; Figure 3B). However, this was very modest and difficult to interpret in view of the slight increase in HIF-2 α protein in the -1 α KO cells.

Ablation of either HIF- α isoform led to significant reduction of HIF-1 β signal across all three sets of binding sites ($p < 0.001$; Figure 3A-F). This is consistent with the inability of one HIF- α isoform to compensate (at least fully) for the absence of the other at any of the sets of sites. This finding also suggests that HIF-1 β is in relative excess, such that its binding is limited by the availability of a HIF- α dimerisation partner.

In order to identify isolated examples of HIF- α binding site alteration in HIF- α KO cells, I performed differential binding analysis using the DiffBind package in R (Ross-Innes et al., 2012). Consistent with previous results, 410 out of the 873 regions bound by both HIF-1 and HIF-2 in WT cells had significantly decreased HIF-1 α binding in the -2 α KO cells using a threshold FDR of 0.05, whilst none had significant increases in binding. ChIP-seq tracks for a representative locus are illustrated in Figure 4A.

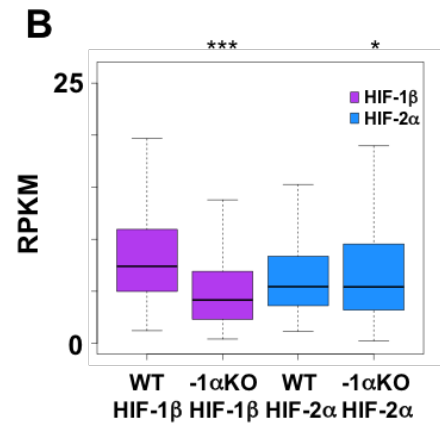
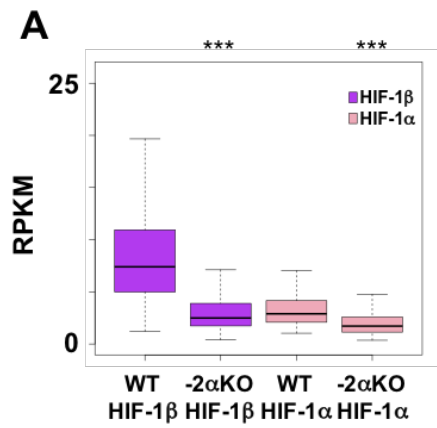
191 of these same regions showed changes in HIF-2 α signal in the -1 α KO cells using the same statistical threshold. However, similar numbers of sites showed an increase (90 regions) or a decrease (101 regions) in signal, consistent with no overall pattern. ChIP-seq tracks for a representative locus are illustrated in Figure 4A.

Whilst there may be a very small number proportion of sites whereby HIF-2 is differentially bound in response to the loss of HIF-1 α (that may indeed be biologically significant), the overall pattern of HIF-1 and HIF-2 DNA binding following the deletion of one or the other HIF- α isoform is that neither HIF-1 or HIF-2 has the capacity to compensate for the other.

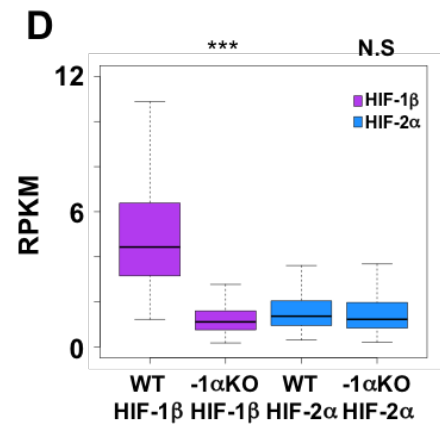
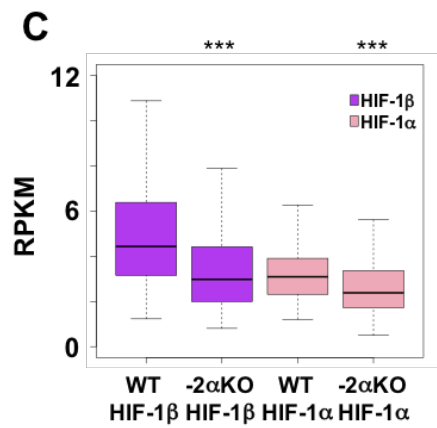
Common HIF-1
and HIF-2 sites

WT vs. -2 α KO

WT vs. -1 α KO



HIF-1 specific sites



HIF-2 specific sites

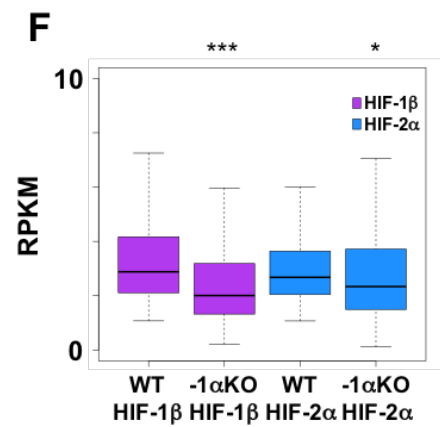
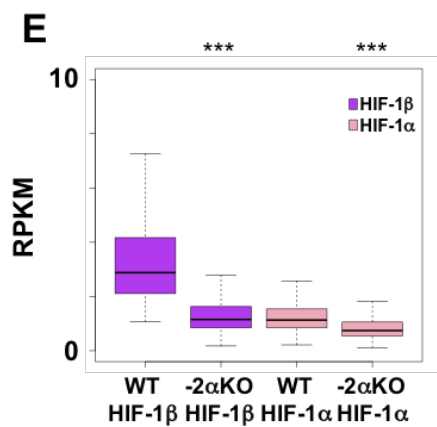


Figure 3. Effects of HIF- α knockout on pan-genomic HIF binding.

(A) Boxplot depicting the read counts per kilobase per million mapped reads (RPKM), across two HIF-1 β (purple) and HIF-1 α (pink) ChIP-seq biological replicates in WT and -2 α KO cells. ChIP-seq signal is being determined within regions bound by both HIF-1 and HIF-2 in HKC8 WT cells. *** $p < 0.001$ compared to RPKM in WT cells. Boxplots show the median and upper and lower quartiles. Whiskers show the interquartile range. Outliers are not shown.

(B) Boxplot depicting the RPKM across two HIF-1 β (purple) and HIF-2 α (pink) ChIP-seq biological replicates in WT and -1 α KO cells. ChIP-seq signal is being determined within regions bound by both HIF-1 and HIF-2 in HKC8 WT cells. *** $p < 0.001$; * $p < 0.05$ compared to RPKM in WT cells.

(C) Boxplot as in (A), however ChIP-seq signal is being determined within regions bound specifically by HIF-1 in HKC8 WT cells. *** $p < 0.001$ compared to RPKM in WT cells.

(D) Boxplot as in (B), however ChIP-seq signal is being determined within regions bound specifically by HIF-1 in HKC8 WT cells. *** $p < 0.001$; N.S. (not significant) compared to RPKM in WT cells.

(E) Boxplot as in (A), however ChIP-seq signal is being determined within regions bound specifically by HIF-2 in HKC8 WT cells. *** $p < 0.001$ compared to RPKM in WT cells.

(F) Boxplot as in (B), however ChIP-seq signal is being determined within regions bound specifically by HIF-2 in HKC8 WT cells. *** $p < 0.001$; * $p < 0.05$ compared to RPKM in WT cells.

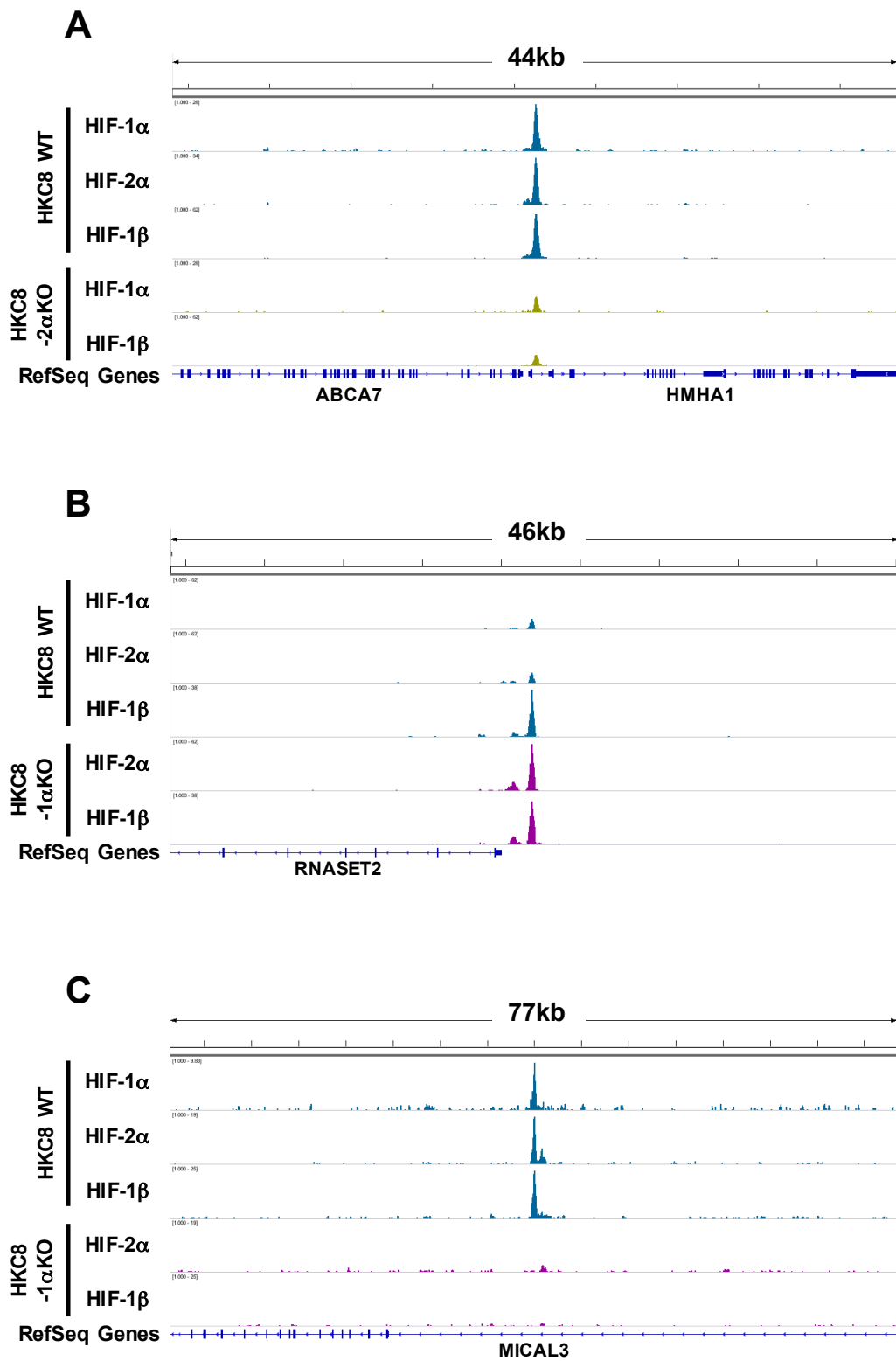


Figure 4. Identification of differentially bound HIF-1 α and HIF-2 α within wild type common HIF-1 and HIF-2 binding sites.

(A) A 44 kb genomic region around the *HMHA1* and *ABCA7* genes illustrating a common HIF-1 and HIF-2 binding site in WT HKC8 that significantly decreases in HIF-1 α binding in -2 α KO cells. Representative HIF-1 α , HIF-2 α and HIF-1 β ChIP-seq samples in WT HKC8 (blue) and -2 α KO HKC8 cells (gold).

(B) A 46 kb genomic region around the *RNASET2* gene illustrating a common HIF-1 and HIF-2 binding site in WT HKC8 that significantly increases in HIF-2 α binding in -1 α KO cells. Representative HIF-1 α , HIF-2 α and HIF-1 β ChIP-seq samples in WT HKC8 (blue) and -1 α KO HKC8 cells (purple).

(C) A 77 kb genomic region within the *MICAL3* gene illustrating a common HIF-1 and HIF-2 binding site in WT HKC8 that significantly decreases in HIF-2 α binding in -1 α KO cells. Representative HIF-1 α , HIF-2 α and HIF-1 β ChIP-seq samples in WT HKC8 (blue) and -1 α KO HKC8 cells (purple).

4.2.4 Functional characterisation of shared HIF-1 and HIF-2 sites

In addition to binding site selectivity, post-binding mechanisms may also contribute to differences in the transcriptional repertoire of the HIF-1 and HIF-2 isoforms (Hu et al., 2003; Raval et al., 2005). To date, the mechanisms governing HIF-1 versus HIF-2 transcriptional selectivity post-binding are poorly understood. In a recent study of genes flanking common HIF-1 and HIF-2 binding sites (Platt et al., 2016), distance from the HIF-binding site to the gene promoter was seen as a factor influencing which isoform regulated the gene. Specifically, gene promoters closer to the HIF-binding sites were more likely to be regulated by HIF-1, whilst those more distant from the HIF binding site were more likely to be regulated by the HIF-2 isoform. Coupled with the differences in binding site distribution, this would enhance the promoter-proximal nature of HIF-1-dependent gene regulation and the promoter-distal nature of HIF-2-dependent gene regulation.

The HKC8 ChIP-seq and RNA-seq datasets were used to explore this phenomenon further. I therefore intersected HIF-1 sites with HIF-2 sites using BEDTools intersect (v2.26.0) in order to identify all sites that were co-occupied by HIF-1 and HIF-2 (Figure 5A). These 872 sites were subsequently annotated by their nearest protein coding gene. All binding sites whose nearest gene was not significantly regulated by hypoxia regulated in WT cells, in addition to being hypoxia regulated in either -1 α KO or -2 α KO cells, were discarded (according to an adjusted *p*-value significance threshold of < 0.05). This resulted in the retrieval of 180 sites, and therefore 180 neighbouring hypoxia-regulated genes.

The 180 shared HIF-1 and HIF-2 sites that neighbour a hypoxia regulated gene are not necessarily *co-occupied* by HIF-1 and HIF-2. These 180 sites may represent sites that are capable of binding both HIF- α isoforms, but are not necessarily bound simultaneously by HIF-1 and HIF-2. Cell lines are a heterogeneous population of cells, therefore it is possible that HIF-1 is bound at these loci in a subset of cells within the population, whilst HIF-2 is bound in another. It is also possible that HIF-1 and HIF-2 are binding interchangeably within the same sub-population of cells at the shared sites in question. Determining whether HIF-1 and HIF-2 are in fact co-occupying these shared loci would be determinable through the use of ChIP re-ChIP, however this is a notoriously noisy technique and was therefore not attempted (Furlan-Magaril et al., 2009). However, given that both HIF-1 and HIF-2 bind HREs within DNA, a greater number of HREs per site may provide an indication of the likelihood HIF-1 and HIF-2 actually co-occupying these loci, rather HIF-1 and HIF-2 binding these sites in a distinct manner in distinct sub-populations of cells.

I therefore quantified the number of HREs per site within the 180 overlapping HIF-1 and HIF-2 sites. I also quantified the number of HREs per site within a random selection of 180 HKC8 WT HIF-1 β sites of equal width to the 180 overlapping HIF-1 and HIF-2 sites. The quantification of HRE density was plotted in the form of a boxplot (Figure 5B). The median number of HREs for both sets of 180 sites is 1, therefore the greatest proportion of sites are likely only bound by one HIF isoform at any one time. However, the mean frequency of HREs in the 180 random HIF-1 β sites is 1.30 per binding site, whereas the mean frequency of HREs in the 180 overlapping HIF-1 and HIF-2 sites is 1.78. Thus suggesting that there may be a greater proportion of the shared HIF-1 and HIF-2 sites that are in

fact co-occupied relative to a random set of 180 HIF-1 β binding sites. A caveat to be noted is that to be included in the set of 180 shared HIF-1 β sites, a site must be present in six ChIP-seq datasets (2 replicates of HIF-1 β , 2 replicates of HIF-1 α and 2 replicates of HIF-2 α). However, to be included within the set of 180 random HIF-1 β sites, a site only needs to be present in two ChIP-seq datasets (2 replicates of HIF-1 β). This means that the 180 shared HIF binding sites may represent higher confidence HIF sites, whereas the 180 random HIF-1 β sites may contain false-positive or noisy sites, hence why the density of HREs within the 180 shared HIF-1 and HIF-2 sites are greater.

To begin to understand how the distance between the shared binding site and the hypoxia regulated gene associates with the binding strength of the HIF- α subunit, and subsequently whether these binding characteristics predict whether the nearby gene will be dependent upon HIF-1 α , HIF-2 α or either HIF- α isoform for its hypoxic regulation, I firstly sorted all 180 shared binding sites by the HIF-1 α to HIF-2 α ChIP-seq signal ratio within these sites (x-axis) (Figure 5C). I subsequently defined each sorted site as to whether the nearby gene was exclusively regulated by HIF-1 α , HIF-2 α , or capable of being regulated by either HIF- α subunit (y-axis). The status of each site with respect to the nearest gene being exclusively regulated by HIF-1 α , HIF-2 α , or capable of being regulated by either HIF- α subunit is indicated by the light pink line. The dark pink line is a moving average of the light pink line. The plot indicates that the greater the HIF-1 α signal relative to HIF-2 α at shared loci, the greater the likelihood that neighbouring gene is regulated exclusively by HIF-1 α . Conversely, the greater

the HIF-2 α signal relative to HIF-1 α at shared loci, the greater the likelihood that neighbouring gene is regulated exclusively by HIF-2 α .

Still with respect to Figure 5C, the distance between the HIF-1 and HIF-2 shared loci and the neighbouring hypoxia-regulated gene was determined for the binding sites sorted by HIF-1 α to HIF-2 α signal ratio. $-\log_{10}[\text{distance in bp}]$ was quantified for each previously determined distance (y-axis), and subsequently plotted (light purple) along the same sorted sites (x-axis) as previously described. The dark purple line is a moving average of the distance in bp between each shared site and its neighbouring hypoxia-regulated gene. The plot indicates that the smaller the distance between a HIF-1 and HIF-2 shared site and the nearest hypoxia regulated gene, the greater the likelihood that the neighbouring gene is regulated exclusively by HIF-1 α . The converse is true with respect to genes exclusively regulated by HIF-2 α .

Three representative genome browser tracks are displayed in Figures 6A-C that illustrate three HIF-1 and HIF-2 co-occupied loci that each neighbour a hypoxia-regulated gene. *SPAG4* represents a gene that is exclusively regulated by HIF-1. *BHLHE40* represents a gene that is capable of being regulated by either HIF- α subunit. And, *ARRDC3* represents a gene that is exclusively regulated by HIF-2.

Taken together, these data indicate that at shared binding sites, the greater the HIF-1 α ChIP-seq signal relative to HIF-2 α signal, the more likely the site is to be promoter-proximal, and the more likely the neighbouring hypoxia-regulated gene is to be exclusively regulated by HIF-1. Conversely, the greater the HIF-2 α ChIP-seq signal relative to HIF-1 α signal at a shared binding site, the

more likely the site is to be promoter-distal, and the more likely the neighbouring hypoxia regulated gene is to be exclusively regulated by HIF-2.

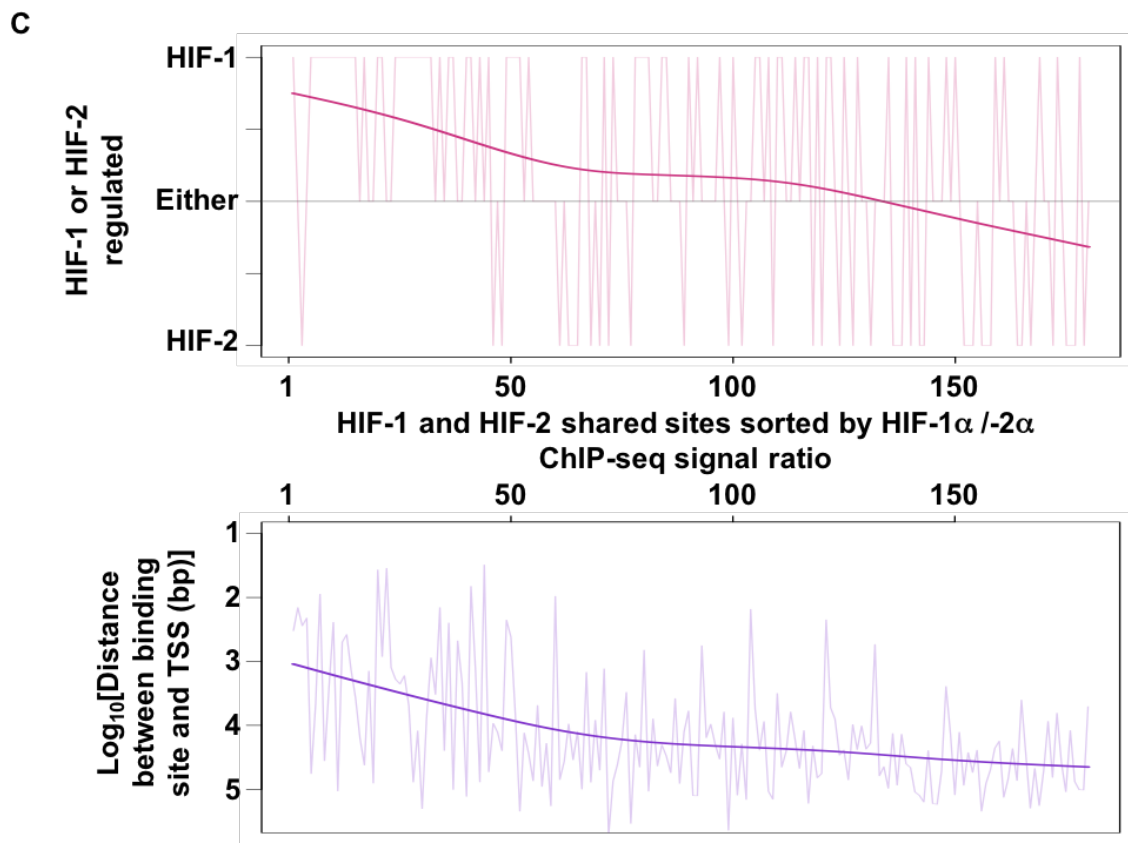
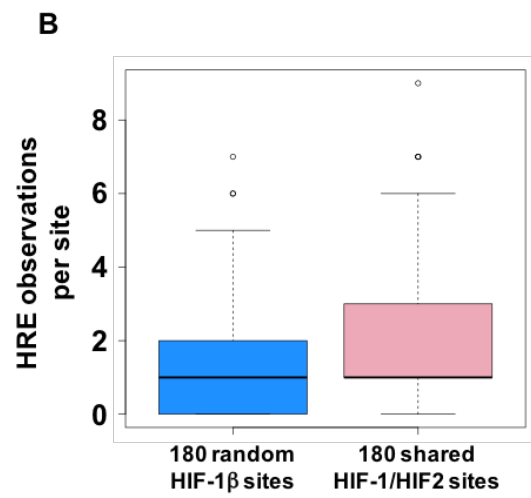
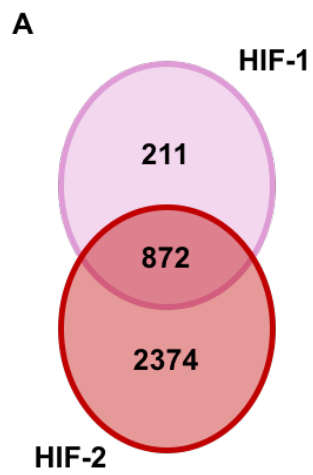


Figure 5. Functional characterisation of HIF-1 and HIF-2 shared binding sites in HKC8 cells

(A) Venn diagram illustrating the number of co-occupied and distinct canonical HIF-1 and HIF-2 sites in HKC8 cells, as determined via biological replicate ChIP-seq assays.

(B) Boxplot illustrating the frequency of HREs (RCGTG) per site. 180 randomly selected HKC8 HIF-1 β sites were compared to the 180 HIF-1 and HIF-2 shared sites that neighbour a hypoxia regulated gene in WT HKC8 and -1 α KO or -2 α KO cells. Equal peak width was ensured by adding 150bp to 5' and 3' of the peak summit coordinates.

(C) The 180 shared HIF-1 and HIF-2 sites assayed previously are sorted along the x-axis of both plots (top and bottom) by the ratio of HIF-1 α to HIF-2 α ChIP-seq signal ratio within these sites. The thick pink line in the top plot represents a moving average of the light pink line, illustrating whether the site neighbouring a gene exclusively regulated by HIF-1, HIF-2 or by either. The thick purple line in the bottom plot represents a moving average of the distance between the shared binding site and the nearest gene.

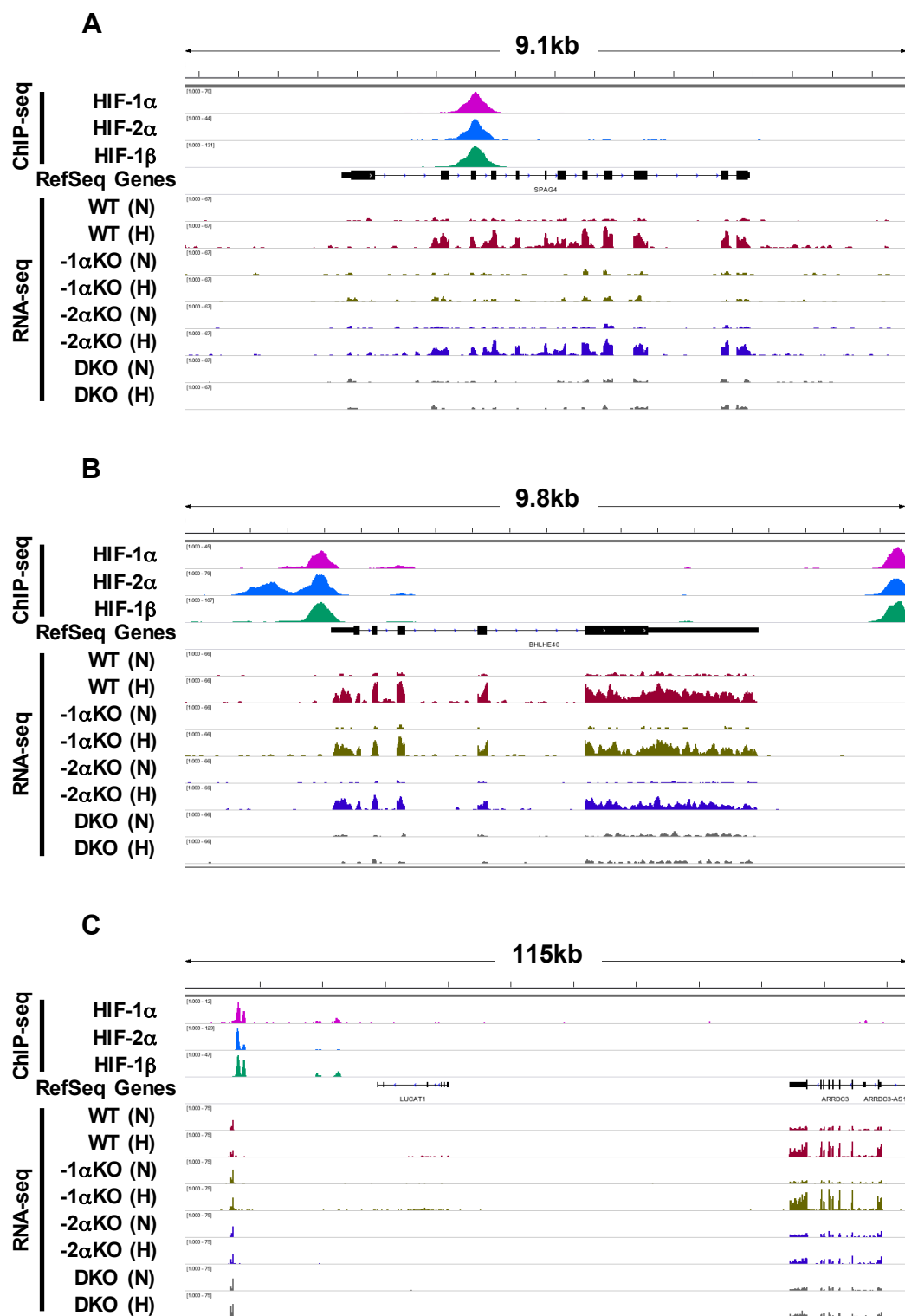


Figure 6. Tracks depicting HIF-1 and HIF-2 shared sites, and the mode of regulation of the nearby gene.

(A) The *SPAG4* gene illustrating a co-occupied HIF-1/2 site and HIF-1 α dependent hypoxia-inducible gene expression. Representative HIF-1 α (pink), HIF-2 α (blue) and HIF-1 β (green) ChIP-seq samples in WT HKC8 cells shown. Representative normoxic (N) and hypoxic (H) (16 hours; 0.5% O₂) RNA-seq samples from WT (red), -1 α KO (green), -2 α KO (blue) and HIF- α double knockout (DKO; grey) cells.

(B) The *BHLHE40* gene illustrating a co-occupied HIF-1/2 site and hypoxia-inducible expression indiscriminate of the HIF- α isoform present. Representative HIF-1 α (pink), HIF-2 α (blue) and HIF-1 β (green) ChIP-seq samples in WT HKC8 cells shown. Representative normoxic (N) and hypoxic (H) (16 hours; 0.5% O₂) RNA-seq samples from WT (red), -1 α KO (green), -2 α KO (blue) and HIF- α double knockout (DKO; grey) cells.

(C) The *ARRDC3* gene illustrating a co-occupied HIF-1/2 site and HIF-2 α dependent hypoxia-inducible gene expression. Representative HIF-1 α (pink), HIF-2 α (blue) and HIF-1 β (green) ChIP-seq samples in WT HKC8 cells shown. Representative normoxic (N) and hypoxic (H) (16 hours; 0.5% O₂) RNA-seq samples from WT (red), -1 α KO (green), -2 α KO (blue) and HIF- α double knockout (DKO; grey) cells.

4.3 Discussion

HIF-1 and HIF-2 both bind DNA via the same consensus motif (RCGTG), otherwise known as the hypoxia response element (HRE) (Wenger et al., 2005). In spite of the shared consensus motif, HIF-1 and HIF-2 bind not only common, but also distinct sites (Mole et al., 2009; Schodel et al., 2011). Published studies have illustrated that HIF-1 and HIF-2 binding sites have distinct pan-genomic distributions with respect to the distance between their binding sites the nearest transcriptional start site (Schodel et al., 2011; Salama et al., 2015). Specifically, HIF-1 sites have a tendency to distribute in a promoter-proximal manner, whereas HIF-2 sites tend to distribute in a promoter-distal manner. In Chapter 3 I demonstrated that sites preferentially bound by either HIF-1 or HIF-2 consistently adopted promoter-proximal or promoter-distal distributions, respectively, across a panel of cell types, despite the binding sites themselves differing greatly. Taken together, this raises questions as to whether HIF-1 and HIF-2 binding sites remain distinct or whether either isoform has the capacity to redistribute in order to compensate for the absence of the other?

In this chapter, CRISPR/Cas9 technology was used to generate HKC8 cell lines that are either simultaneously, or separately, deficient in HIF-1 α and HIF-2 α . I then performed HIF ChIP-seq across the WT and mutant cell lines to illustrate the extent to which HIF-1 α and HIF-2 α can exhibit compensation or binding site interchangeability. Furthermore, I utilised these cellular reagents in order to shed further light on the functional consequences of sites shared by HIF-1 and HIF-2.

4.3.1 HIF does not undertake widespread redistribution to compensate for the absence of the other HIF- α isoform

In this chapter, I performed pan-genomic analyses of HIF-1 β , HIF-1 α and HIF-2 α chromatin-immunoprecipitates from HKC8 WT cells, in addition HKC8 cells that were engineered by CRISPR/Cas9 to independently harbour inactivating mutations in *HIF1A* and *EPAS1* genes. Performing HIF ChIP-seq in the HIF- α knockout and WT settings enabled me to compare and contrast HIF binding in the WT and mutant settings.

Four different analysis methodologies were implemented in order to determine the capacity of HIF-1 α and HIF-2 α to redistribute in the HIF- α KO cells relative to the WT. Whether analysed qualitatively via peak calling, or quantitatively by measuring ChIP-seq RPKM in defined regions, HIF-1 α and HIF-2 α appeared largely incapable of undertaking a binding redistribution to compensate for the absence of the other HIF- α isoform. However, HIF-2 α exhibits an exception to this overall trend by undertaking a modest redistribution from regions bound specifically by HIF-2 in WT cells, to sites commonly bound by HIF-1 and HIF-2, following the knockout of HIF-1 α . Reassuringly, the sites whereby HIF-2 α binding increases in -1 α KO cells remain associated with hypoxia inducible genes in -1 α KO cells, thus suggesting that HIF-2 α binding is sufficient to maintain hypoxia-inducibility of these genes in the absence of HIF-1 α .

In the absence of HIF-2 α , HIF-1 α exhibited a quantitative decrease in DNA binding within all assessed sites. This may be a consequence of the observed decrease in HIF-1 α protein levels in -2 α KO cells relative to WT. The decrease in HIF-1 α protein may be considered surprising given published reports

demonstrating that in hepatocellular carcinoma cells, stable knockdown of either HIF-1 α or HIF-2 α leads to enhanced protein abundance of the other HIF- α subunit, thus leading to a survival advantage in hepatocellular carcinoma development (Menrad et al., 2010; Schulz et al., 2012). However, a well known caveat to generating a CRISPR/Cas9 engineered cell line are single-cell clonal effects that manifest themselves as a result of the initial cell being derived from a heterogeneous aneuploid population of cells. To sufficiently overcome the single-cell clone caveat, it is necessary to subject multiple validated knockout clones to the same experimental assays. For the entirety of this chapter, I only used one clone for all experiments, therefore future work should utilise additional clones when attempting to validate the findings described in this chapter.

My data has demonstrated a remarkable rigidity of HIF-1 and HIF-2 in their relative mechanisms of binding site selection. In the absence of the other HIF- α isoform, neither HIF-1 α or HIF-2 α undergoes a significant degree binding redistribution. In relation to this finding is a published study whereby overexpression of exogenous HIF-1 α and HIF-2 α in renal cancer cells still leads to both factors adopting their respective typical promoter-proximal and promoter-distal distributions (Salama et al., 2015). The implications of these findings strongly suggest that there are specific factors that are uniquely interacting with HIF-1 α or HIF-2 α to determine their spectrum of binding sites within chromatinised DNA. It is possible that these interacting factors may be listed among the ENCODE transcription factor ChIP-seq datasets that significantly coincided with either HIF-1 or HIF-2 binding sites, as described during Chapter 3.

4.3.2 The functional consequences of HIF-1 and HIF-2 shared binding sites

By combining ChIP-seq analyses with transcript profiling in normoxia and hypoxia in WT and HIF- α mutant cells, I was able to gain an understanding of the functional consequences of HIF-1 and HIF-2 binding to co-occupied loci. I determined that the greater the HIF-1 α ChIP-seq signal relative to HIF-2 α signal at a shared binding site, the more likely the site is to be promoter-proximal, and the more likely the neighbouring hypoxia-regulated gene is to be exclusively regulated by HIF-1. The inverse of these characteristics increases the likelihood of the neighbouring hypoxia-regulated gene to the co-occupied HIF binding sites being exclusively regulated by HIF-2. The association that I have drawn between distance between a shared site and a hypoxia regulated gene corroborates initial findings introduced in an earlier study (Platt et al., 2016).

A fundamental caveat to consider here is that the cell line being used (HKC8 WT) is a heterogeneous population of aneuploid cells, meaning that the same site may be behaving differently within each cell in the population. Of concern is that fact HIF-1 and HIF-2 may not be co-occupying a site, rather more HIF-1 may be binding in a subset of the population, whereas HIF-2 may be binding in another. However, the increased proportion of HREs present within these sets defined as shared sites suggests that there is an increased likelihood of HIF-1 and HIF-2 binding locally to one another.

Taken together, these data illustrate that binding distributions of HIF-1 and HIF-2 are remarkably rigid within a given cell type. Moreover, in circumstances whereby HIF-1 and HIF-2 are binding common sites, it should not be assumed that both HIF-1 and HIF-2 are capable of transactivating the expression of the nearby gene in hypoxic conditions. These are considerations that should be

considered during the development of HIF- α -specific antagonists (Wallace et al., 2016). Specifically, HIF-1 α or HIF-2 α therapeutics intervention is unlikely to evoke widespread compensatory activities by the other HIF- α isoform, however the remaining present HIF- α isoform will remain capable of maintaining robust transactivation of its spectrum of target genes.

Chapter 5: Cooperativity between HIF and AP-1

5.1 Introduction

As was described in Chapter 3, I demonstrated that the AP-1 consensus motif, otherwise known as the 12-O-Tetradecanoylphorbol-13-acetate (TPA) response element (TRE), was significantly enriched among regions preferentially bound by HIF-2. Moreover, I demonstrated that HIF-2 binding sites, as determined via ChIP-seq performed in hypoxia in HepG2 cells, showed a highly significant coincidence with the AP-1 family member binding sites, as determined via the ENCODE Consortium performing ChIP-seq in HepG2 cells. Taken together, these data suggested that HIF-2 and AP-1 are selectively co-occupying *cis*-regulatory elements, and may therefore be engaging in functional cooperativity.

AP-1 is a dimeric basic leucine zipper transcription factor that may exhibit both basal and inducible transcriptional activity via the engagement of the TRE consensus sequence in DNA (TGAG/CTCA) (Hess et al., 2004). Whilst there are numerous families of structurally related proteins that may dimerise to form a functional AP-1 transcription factor, the families that can dimerise and engage a TRE are the JUN (JUN, JUNB, JUND) and FOS (FOS, FOSB, FOSL1, FOSL2) families (Chinenov and Kerppola, 2001). JUN family members may dimerise with a fellow member of the JUN family, or heterodimerise with a FOS family member. However, FOS family members may not homodimerise within the FOS family, they may only heterodimerise with the JUN family.

AP-1 is a pleiotropic transcription factor that may be activated by a broad range of physiological and pathophysiological stimuli, such as growth factors, cytokines, stress signals, viral and bacterial infections, in addition to oncogenic stimuli (Chang and Karin, 2001; Hill et al., 1994). Specifically, with regards to

tumourigenesis, reports have illustrated that the composition of the AP-1 dimer can result in a pro- or anti-tumourigenic entity (reviewed here: Shaulian, 2010).

Numerous studies have demonstrated a functional association between AP-1 and the HIF transcription factors (Laderoute, 2005). Specifically, many of these studies have demonstrated the capacity of AP-1 to potentiate the expression of a HIF target genes in hypoxia, such as *VEGFA* (Damert et al., 1997), *FGF2* (Lee and Corry, 1999), *CXCL8* (Desbaillets et al., 1999) and *EDN1* (Yamashita et al., 2001), amongst others. The mechanisms relating to these HIF and AP-1 gene regulatory events will be further discussed later in this chapter with respect to the findings I have uncovered relating to the pan-genomic relationship between HIF and AP-1. Another study predicted HIF and AP-1 binding site co-occupancy, however the authors subsequently demonstrated a slightly enhanced capacity for hypoxic induction of *CA9* upon mutation of an AP-1 motif neighbouring the HRE in *CA9* promoter (Villar et al., 2012), thus suggesting that AP-1 may be augmenting the hypoxic inducibility of some HIF target genes by suppressing the ability of HIF to fully transactivate target gene expression in hypoxia. A more recent pan-genomic HIF binding study also predicted HIF and AP-1 binding site co-occupancy (Salama et al., 2015). Specifically, upon overexpression of HIF-2 α in 786O cells, a highly significant TRE motif enrichment was observed, thus suggesting that HIF-2 α may have the capacity to co-occupy a large number of *cis*-regulatory elements with AP-1.

The primary aim of this chapter is to explore potential pan-genomic transcription factor cooperativity between HIF-2 and AP-1, whereby AP-1 is hypothesised to be facilitating the DNA binding of HIF-2 and consequently enabling HIF-2 to transactivate expression of its target genes. Providing a

framework for this investigation are studies that have utilised pan-genomic technologies to elucidate novel interactions between other transcription factors. Jason Carroll's laboratory, one of the pioneering laboratories for ChIP-seq studies, made a groundbreaking discovery that the forkhead box (FOX) transcription factor motif was significantly enriched within estrogen receptor binding sites relative to random background genomic regions, and that the FOXA1 transcription factor was required for estrogen receptor (ER) to bind to the entirety of its target *cis*-regulatory elements in ER+ breast cancer cells (Carroll et al., 2005; Laganière et al., 2005). AP-1 binding has been shown to increase chromatin accessibility, thus enabling the glucocorticoid receptor (GR) to bind approximately half of its target *cis*-regulatory elements in mammary epithelial cells (Biddie et al., 2011). A similar mechanism is mediated by the C/EBP in hepatic cells in order to facilitate the binding of the GR to a significant proportion of its target loci (Grontved et al., 2013). More recently, AP-1 has been shown to cooperate with the TEAD family of transcription factors, whereby AP-1 facilitates the binding of TEAD to active *cis*-regulatory elements (Zanconato et al., 2015), whilst cooperating in the activation of a transcriptional programme to evoke cell migration and invasion (Liu et al., 2016).

Taken together, the framework provided by these studies – particularly those that highlight the role of AP-1 in transcriptional cooperativity, in conjunction with my data outlined in Chapter 3, led me to postulate that HIF and AP-1 might be engaging in DNA-binding cooperativity. No study has yet demonstrated whether HIF and AP-1 exhibit any dependence on one another for their binding to chromatinised DNA.

One hypothesis underlying the work described in this chapter was that AP-1 engages in preferential cooperativity with HIF-2 as opposed to HIF-1. I therefore aimed to investigate the potential relationship between HIF and AP-1 through the utilisation of an intervention against AP-1 that ablates all pan-genomic DNA binding of AP-1, thus enabling a subsequent assessment of the downstream effects upon the pan-genomic binding of HIF.

5.2 Results

5.2.1 AP-1 intervention utilising the AP-1 dominant-negative, AFos

To determine the influence of AP-1 on HIF binding site selectivity, I devised an intervention strategy to ablate AP-1 DNA binding and assessed the results by pan-genomic analyses on HIF binding and hypoxia-inducible gene expression. It was argued that if HIF was dependent on AP-1 for binding at any of its target loci, I would be able to observe this dependency by HIF ChIP-seq experiments in the presence and absence of the intervention.

The intervention that I opted to utilise was an AP-1 dominant-negative that was generated by Charles Vinson's lab in 1997 (Olive et al., 1997). The dominant-negative, AFos (Acidic Fos), comprises an acidic amphipathic C-terminus that replaces the usual basic DNA binding domain, appended to the murine Fos dimerisation domain. The AFos sequence is also appended to a FLAG sequence to enable ease of detection and pull-down during experimental utilisation. The resulting dominant-negative therefore functions via dimerisation with JUN family members, thus preventing their JUN-JUN homodimerisation or JUN-FOS heterodimerisation, and subsequently preventing any functional AP-1 dimer engagement of the consensus TRE. The lack of a basic DNA binding domain in AFos means that AFos-JUN dimers are not capable of binding DNA. This mechanism of action has been depicted by a schematic shown in Figure 1A.

I opted to introduce AFos into the clear cell renal carcinoma cell line, RCC4. A renal cell was chosen because in Chapter 3 I showed that the AP-1 consensus motif has been observed to be more significantly enriched within HIF-2 binding sites in renal cells (HKC8, RCC4), relative to HIF-2 binding sites in HepG2. Furthermore, AP-1 has been shown to be functionally significant for renal

carcinoma tumourigenesis. Specifically, JUNB has been shown to promote cell invasion and angiogenesis in a ccRCC (clear cell renal cell carcinoma) xenograft model (Kanno et al., 2012). Similarly, HIF-2 α has been shown to promote tumour progression when overexpressed in renal carcinoma xenograft assays (Raval et al., 2005). I therefore postulated that a cooperative genome-wide DNA-binding relationship between HIF-2 and AP-1 may be enabling the HIF pathway in renal carcinoma to be more oncogenic.

RCC4 cells were specifically chosen among the numerous ccRCC lines because they express both HIF-1 α and HIF-2 α . The presence of both HIF-1 α and HIF-2 α in the cell line of choice is necessary to test my hypothesis that AP-1 is functionally cooperating with HIF-2 and not HIF-1. This was based on the findings outlined in Chapter 3 that the TRE preferentially enriches within HIF-2 sites, and HIF-2 binding sites significantly overlapped the AP-1 binding sites derived from the ENCODE ChIP-seq datasets.

The method used to introduce AFos into RCC4 cells was lentiviral transduction. The FLAG-tagged AFos nucleotide sequence was ligated into the tetracycline-inducible (Tet-On) lentiviral expression vector, pTRIBZ. Lentivirus containing pTRIBZ-AFos was made by transfecting 293T cells with pTRIBZ-AFos, along with lentiviral packaging (pCMV- Δ 8.91) and envelope vectors (pCMV-VSV-G). 48 hours after transfection, the 293T culture media was sterile filtered and applied to RCC4 cells. Positively transduced cells were selected by blasticidin treatment 48 hours after lentiviral transduction.

RCC4 cells transduced with pTRIBZ Vector alone (-VA) or pTRIBZ-AFos were assessed for FLAG-AFos expression via western blot analysis. Cells were treated with vehicle or doxycycline (1 μ g/ml) for 48 hours prior to harvesting cells

for their whole cell extract and subsequently probing for FLAG-AFos via western blot (Figure 1B). Reassuringly, no FLAG-AFos was detected in RCC4 cells transduced with pTRIBZ-VA. Likewise, no FLAG-AFos was detected in vehicle-treated RCC4 pTRIBZ-AFos cells, thus indicating that there is little or no 'leaky' expression of AFos in the pTRIBZ Tet-On system. Importantly, strong induction of FLAG-AFos is observed in RCC4 pTRIBZ-AFos cells following 48 hours of doxycycline treatment.

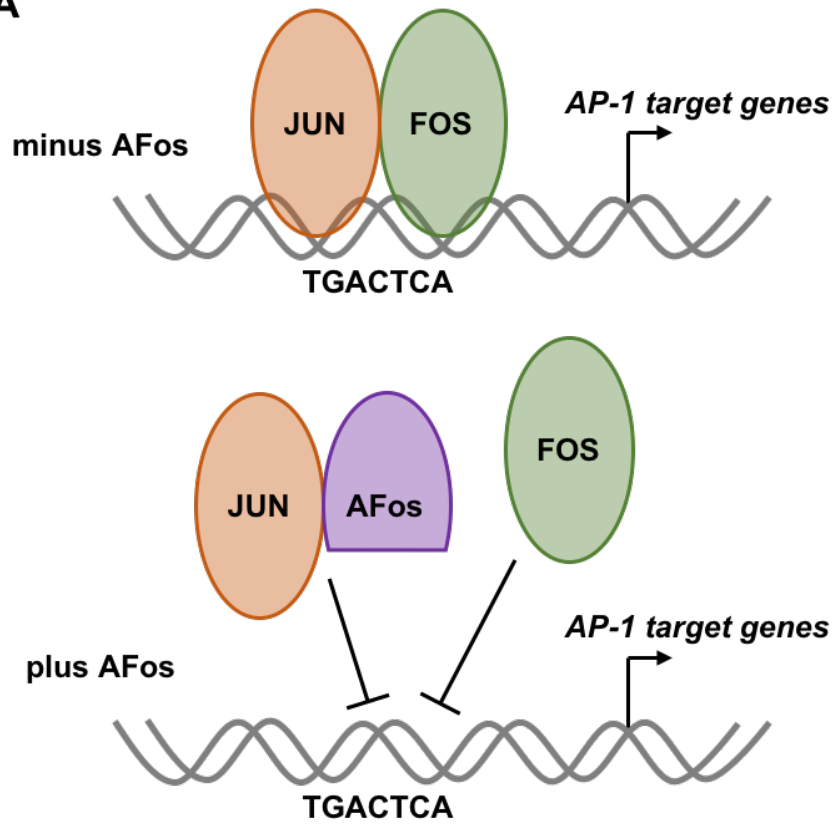
To determine the functional efficacy of AFos in abrogating the binding of AP-1 to DNA, ChIP-qPCR was performed on nuclei harvested from RCC4 pTRIBZ-AFos cells treated with vehicle or doxycycline for 48 hours. JUND was chosen as a surrogate for AP-1 binding because it is highly expressed in RCC4 cells and a high-fidelity ChIP-grade antibody is available. Enrichment of JUND binding was tested using oligonucleotides designed against a known AP-1 binding site upstream of the *SEMA4B* gene. Three biological replicate experiments were performed and a mean JUND binding reduction of 94% was observed when cells were treated with doxycycline (** $p < 0.001$; Figure 1C).

To rule out the possibility of the reduction in JUND enrichment at the *SEMA4B* locus being a result of AFos acting as a blocking peptide against the JUND ChIP antibody, I performed an amino acid sequence alignment of the antigen sequence of the JUND antibody and the amino acid sequence of AFos (Figure 1D). Clustal Omega Multiple Sequence Alignment was used for amino acid sequence alignment (Sievers et al., 2014). No sequence alignment was observed.

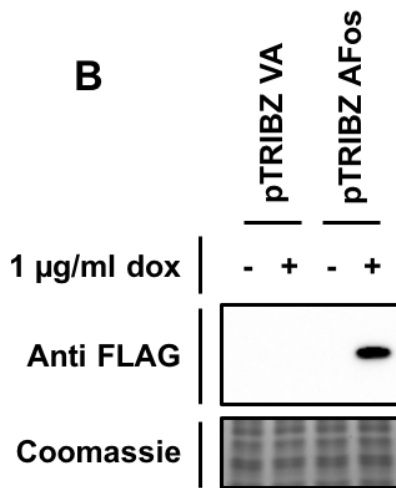
Taken together, the control sequence alignment measures taken to ensure AFos is not acting as a blocking peptide, in addition to the ChIP-qPCR data and

non-leaky expression of AFos in the Tet-On system, suggest that AFos should provide a valid intervention strategy to ablate AP-1 binding to DNA and assess the consequences across the genome.

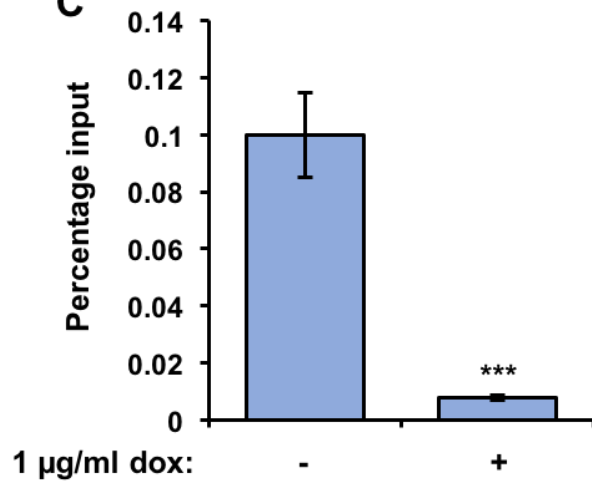
A



B



C



D

FLAG-AFos	MDYKDDDDKHMASMTGGQQMGRDPDLEQRAEELARENEELEKEAEELEQELAELEAETDQ
JUND Ab epitope	-----AQ
	*
FLAG-AFos	LEDEKSALQTEIANLLKEKEKLEFILAAHRPACKIPD-----
JUND Ab epitope	L-----KQKVLSHVNSGCQLLPQHQPAY
	* : * : . * : :

Figure 1. Validation of the doxycycline inducibility of the AP-1 dominant-negative, AFos and its ability to block JUND binding to chromatin.

(A) Schematic depicting the mechanism of action of AFos. Saturating levels of AFos competitively block the ability of JUN to dimerise with other JUN and FOS family members. AFos containing heterodimers are unable to bind DNA due to the substitution of the basic domain for acidic residues. Consequently, AP-1 cannot bind the TRE consensus element and transactivate AP-1 target genes.

(B) Western blot of FLAG expression in RCC4 cell lysates stably transduced with lentiviral pTRIBZ empty vector (VA) or pTRIBZ AFos. Cells were treated with vehicle or with 1 µg/ml doxycycline for 48 hours.

(C) ChIP-qPCR performed using an antibody against JUND antibody in the presence of vehicle or 1 µg/ml doxycycline for 48 hours. Enrichment of JUND binding at the *SEMA4B* enhancer region is shown, as a percentage of input chromatin. Error bars represent $\pm$ the standard deviation of the percentage input across three biological replicate experiments.

(D) Amino acid sequence alignment of FLAG-AFos and the JUND ChIP antibody's target epitope within the JUND amino acid sequence.

5.2.2 Pan-genomic validation of AFos

Biological duplicate ChIP-seq experiments were performed using antibodies directed against HIF-1 α , HIF-2 α , HIF-1 β and JUND in RCC4 pTRIBZ-AFos cells treated with either vehicle or doxycycline (1 μ g/ml) for 48 hours and peaks were called using the MACS peak caller as before. Having shown by ChIP-qPCR that AFos abrogates the binding of JUND at a single locus, I first determined whether expression of AFos effectively suppressed JUND binding across the entire genome. JUND ChIP-seq signal (CPM) was averaged across both replicates and all binding regions identified in the MACS peak calling analysis. The average signal was then plotted as a function of the distance either side of the peak summits (Figure 2A). A strong JUND binding signal, centered on the peak summits, was observed in the unstimulated control cells. This signal was largely abrogated in the cells expressing AFos indicating that the dominant negative form of AP-1 had, indeed inhibited pan-genomic binding of at least one AP-1 subunit. ChIP-seq tracks showing biological replicate JUND ChIP-seq signal (red) across a representative region are shown to illustrate this further (Figure 2B). Since AFos will heterodimerise with all JUN family members, the implication is that all AP-1 dimers that bind the TRE, will be similarly inhibited by AFos.

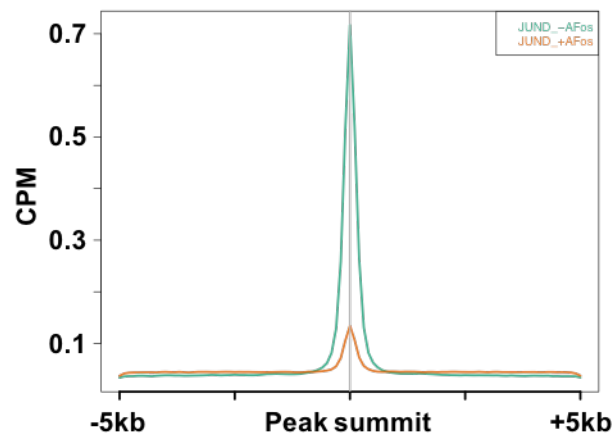
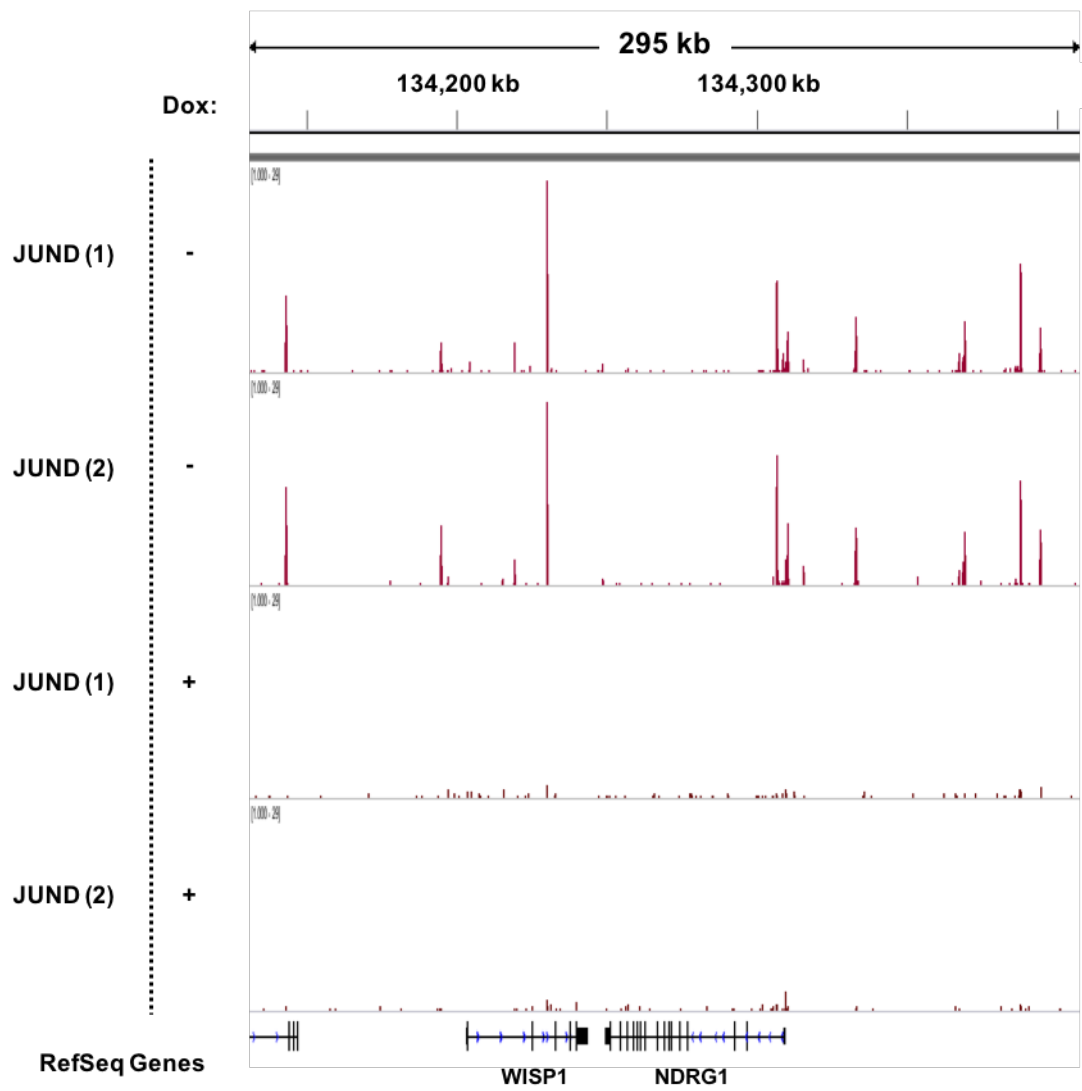
A**B**

Figure 2. Validation of the genome-wide efficacy of AFos by ChIP-seq.

(A) Average normalised JUND ChIP-seq signal (counts per million, CPM) for ± 5 kb regions flanking the summits of peaks identified in the un-induced control cells. Signals were averaged across all identified binding sites and both biological replicates. Signals from un-induced control cells are shown in green and from cells stimulated with doxycycline to express the dominant negative AFos are shown in orange.

(B) IGV tracks for a 295 kb genomic region around the *WISP1* and *NDRG1* genes depicting normalised ChIP-seq signal (CPM) for each biological replicate of the JUND (red) immunoprecipitations in the presence and absence of AFos.

5.2.3 AP-1, its pan-genomic association with HIF, and the effects of AFos on HIF DNA-binding

I next determined the effect of AFos induction, on the pan-genomic HIF-binding signal. Specifically, I examined HIF-1 α , HIF-2 α and HIF-1 β ChIP-seq signal across all regions identified by the MACS peak-calling algorithm as binding HIF-1 β in the vehicle-treated control cells. I chose to examine HIF-1 β peaks since both canonical HIF-1 and canonical HIF-2 sites will contain HIF-1 β binding signals. For each region, the ChIP-seq signal (RPKM) for each JUND, HIF-1 β , HIF-1 α and HIF-2 α ChIP-seq experiment was determined and averaged across the two biological replicates. Results are displayed in the form of boxplots to show the median, interquartile range and full-range of signal for each ChIP-seq experiment (Figure 3A). Inspection of HIF-2 α ChIP-seq in vehicle-treated cells indicated that one biological replicate yielded much lower signal than all other HIF ChIP-seq analyses, including the other biological replicate of the same conditions. This replicate was therefore omitted from all further analyses. Statistically significant differences in signal between ChIP experiments performed in the presence or absence of AFos were determined using an independent 2-group unpaired t-test.

Consistent with the analysis above, a significant decrease in JUND (green) signal was observed within HIF-1 β binding sites following induction of AFos ($p < 2.2\text{e-}16$) and the average JUND signal fell by 73%. HIF-1 β ChIP-seq signal across the HIF-1 β binding sites also fell following induction of AFos (19% reduction, $p = 8.3\text{e-}10$). Similar patterns were observed for HIF-1 α (22% reduction, $p = 6.5\text{e-}12$) and for HIF-2 α (11% reduction, $p = 8.3\text{e-}5$) across these regions following induction of AFos.

Although highly statistically significant, the average reductions in HIF-1 α , HIF-2 α and HIF-1 β signal were less than for JUND. This might be expected, if effects on HIF binding were occurring in *cis*, since they might be restricted to the subset of HIF-1 β binding sites that were also bound by AP-1. Conversely, if effects were occurring in *trans*, then similar effects on HIF-binding might be observed at sites that bound HIF without AP-1 as at sites that bound both transcription factors.

Therefore, I next examined the effects of AFos on HIF-binding sites co-occupied by JUND and HIF-binding sites that did not bind JUND. I first used BEDTools intersect (v2.26.0) to identify all HIF-1 β sites that overlapped a JUND binding site by one or more base pairs. This defined 575 HIF-1 β binding sites that were 'co-occupied' by JUND, and 492 HIF-1 β sites that did not overlap a JUND binding site (Figure 3B).

I then repeated the above analysis on each subset of HIF-1 β binding sites. As before, a comparable decrease in JUND signal was again observed within the 'co-occupied' binding sites following induction of AFos (Figure 3C) (78% fall, $p < 2.2\text{e-}16$). The HIF-binding signal also fell within these 'co-occupied' sites. Notably, both the magnitude of the reduction (HIF-1 α , 30%; HIF-2 α , 25%; and HIF-1 β , 31%) and the statistical significance (HIF-1 α , $p = 2.2\text{e-}16$; HIF-2 α , $p = 3.3\text{e-}12$; HIF-1 β , $p = 2.2\text{e-}16$) were greater than when all HIF-1 β binding sites were considered together.

Conversely, when HIF-1 β sites that did not overlap JUND sites were considered, although JUND binding was much lower, there was still some reduction in JUND signal (55% fall, $p < 2.2\text{e-}16$) when AFos was induced (Figure 3D). Consistent with the arguments outlined earlier in my thesis, this would

suggest that JUND was still binding to at least some of these sites, despite the absence of a MACS defined peak. Nevertheless, I proceeded to analyse HIF-binding at these sites. In contradistinction to the sites that were 'co-occupied' by HIF and JUND, neither HIF-2 α , nor HIF-1 β signal were significantly affected by AFos induction at these sites. Although HIF-1 α binding was significantly reduced, both the magnitude and the statistical significance of this fall (17% fall, $p = 3.0e-4$) were less than for the co-occupied sites. This would suggest that the ability of AP-1 to promote HIF binding is predominantly occurring through effects on the chromatin in *cis* rather than on the HIF transcription factors themselves in *trans*.

However, given that a number of these sites still showed evidence of JUND binding that was being attenuated by AFos, I next used a more quantitative technique, in line with that used in previous chapters to re-define HIF-1 β sites that were either co-occupied by both transcription factors or sites that bound HIF-1 β sites alone. HIF-1 β binding sites were first ranked according to JUND signal. I then stratified the 200 HIF-1 β peaks with the greatest JUND signal, and the 200 peaks with the lowest JUND to define the two subsets of HIF-1 β binding sites (Figure 3E). ChIP-seq signals for JUND and each HIF subunit were then displayed for each subset of sites using boxplots as previously (Figure 3F-G).

This analysis again emphasises the influence of the local JUND signal intensity on the magnitude of the effect of AFos induction on HIF binding. Specifically, at the 200 sites binding JUND most strongly, HIF-1 α ($p < 3.2e-10$), HIF-2 α ($p < 2.5e-8$) and HIF-1 β ($p < 3.4e-12$) binding were again significantly reduced by AFos induction (Figure 3F). Conversely at the 200 HIF-1 β sites with the weakest JUND signal, the effects of AFos induction on HIF-1 α ($p = 0.038$),

HIF-2 α ($p = 0.01$) and HIF-1 β ($p = 0.24$) were far more modest (Figure 3G), again consistent with an effect of AP-1 on HIF binding in *cis* rather than in *trans*.

Importantly, both of these analyses show an effect of AFos on the binding of HIF-1 α and of HIF-2 α indicating that AP-1 promotes the binding of both HIF isoforms. However, given that AP-1 has a promoter-distal distribution that is more similar to that of HIF-2 α than HIF-1 α , AP-1 binding coincides with HIF-2 α binding more frequently than it does with HIF-1 α (62% versus 45%, Figure 3H-I). Therefore AP-1 will have more influence on HIF-2 α binding than on HIF-1 α binding, simply because they bind at the same site more frequently. However, the ability of AP-1 to similarly promote binding of both HIF-1 α and HIF-2 α , suggests that it is not the mechanism, by which HIF-1 α and HIF-2 α are directed to different sites on chromatin.

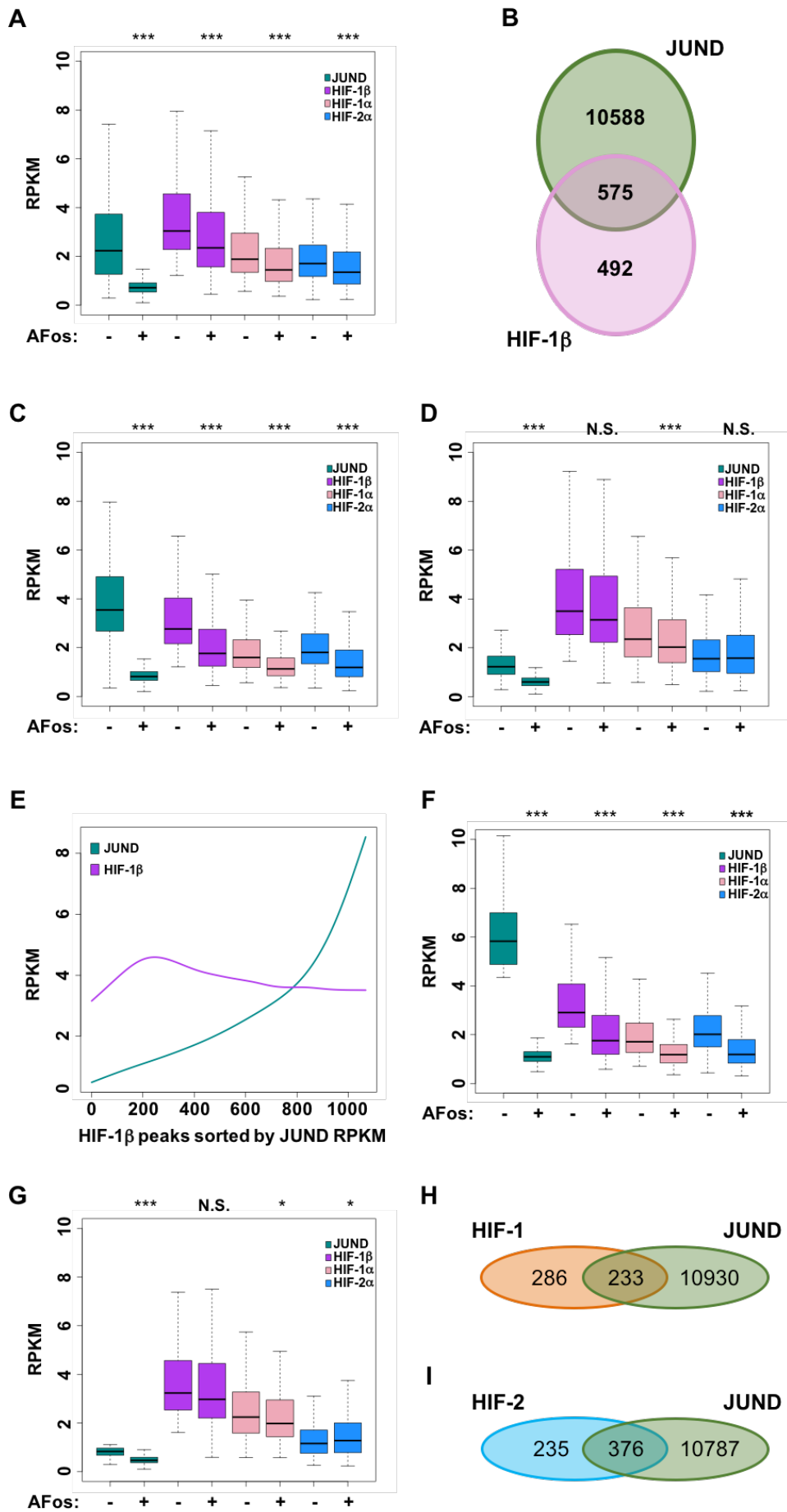


Figure 3. Effects of the loss of JUND binding on HIF binding at sites shared and not shared with JUND.

(A) Boxplot depicting the read counts per kilobase per million mapped reads (RPKM), across two JUND (green), HIF-1 β (purple), HIF-1 α (pink) and HIF-2 α (blue) ChIP-seq biological replicates, within HIF-1 β -bound sites in vehicle-treated RCC4 pTRIBZ AFos cells as compared to cells treated with 1 μ g/ml doxycycline for 48 hours. *** $p < 0.001$ compared to vehicle-treated control samples. Boxplots show the median and upper and lower quartiles. Whiskers show the minimum and maximum values.

(B) Venn diagram illustrating the number of co-occupied and distinct HIF-1 β and JUND binding sites in RCC4 pTRIBZ AFos cells, as determined via biological replicate ChIP-seq assays.

(C) Boxplot as in (A), however RPKM values displayed are calculated within sites bound by HIF-1 β and JUND.

(D) Boxplot as in (A), however RPKM values displayed are calculated within sites bound by HIF-1 β and not JUND.

(E) Line plot depicting the moving average of HIF-1 β RPKM (purple) within HIF-1 β peaks that have been sorted by JUND RPKM, both in the absence of AFos. HIF-1 β peaks with low and high JUND signal may stratified and subjected to further analyses.

(F) Boxplot as in (A), however RPKM values displayed are calculated within the top 200 HIF-1 β peaks with respect to their JUND RPKM.

(G) Boxplot as in (A), however RPKM values displayed are calculated within the bottom 200 HIF-1 β peaks with respect to their JUND RPKM.

(H) Venn diagram illustrating the number of co-occupied and distinct HIF-1 and JUND binding sites in RCC4 pTRIBZ AFos cells, as determined via biological replicate ChIP-seq assays.

(I) Venn diagram illustrating the number of co-occupied and distinct HIF-2 and JUND binding sites in RCC4 pTRIBZ AFos cells, as determined via biological replicate ChIP-seq assays.

5.2.4 Differential binding of HIF in response to AFos

The above analyses indicate average patterns of behavior across all sites co-occupied by HIF and JUND. However, it is possible that AP-1 only influences HIF binding at a subset of these sites. I therefore performed differential binding analysis using the R package, DiffBind (Ross-Innes et al., 2012) to determine which amongst the co-occupied sites were affected by AFos induction. Defining HIF sites that are and are not affected by AFos inductions enables characterisation of HIF sites for subsequent analyses to be outlined in later sections.

Firstly, I assessed HIF-1 β ChIP-seq signal in the presence and absence of AFos, within the 575 regions that are co-occupied by JUND in vehicle-treated RCC4 pTRIBZ-AFos cells. Read counts for each biological replicate ChIP-seq sample were normalised to input control reads. DiffBind quantified mean read counts across the two HIF-1 β ChIP-seq biological replicates in vehicle-treated and doxycycline-treated cells, and determined binding fold change at each site, and differential binding significance.

Of the 575 regions bound by both HIF-1 β and JUND in vehicle-treated cells, 258 were identified as being significantly differentially bound by HIF-1 β following the induction of AFos when using the default threshold FDR of 0.05. 253 sites decreased whereas 5 sites increased in HIF-1 β binding. A representative locus depicting a decrease in HIF-1 β binding in response to AFos has been illustrated by ChIP-seq tracks (Figure 4A).

Of the 492 regions bound by HIF-1 β and not JUND in vehicle-treated cells, 30 were identified as being significantly differentially bound by HIF-1 β following the induction of AFos using the default threshold FDR of 0.05. 27 of the 30

regions decrease in HIF-1 β binding following AFos induction relative to vehicle-treated cells, whereas 3 regions increase in HIF-1 β binding. A representative locus depicting a decrease in HIF-1 β binding in response to AFos has been illustrated by ChIP-seq tracks (Figure 4B).

Note that JUND ChIP-seq signal in the absence of AFos in Figure 4B (red) appears to represent a *bona fide* binding site, however this region was not called as a peak and thus suggests that my earlier postulation regarding MACS under-calling JUND ChIP-seq peaks may be correct. Further exploration of these sites illustrate that JUND signal indicative of a peak is present alongside a HIF site, despite not being called in both replicates by MACS. This suggest MACS has under-called JUND peaks. It also suggests that the effects of AP-1 on HIF binding are in *cis* rather than *trans*.

Taken together, these data show that a significant subset of the HIF-1 β sites that are co-occupied with JUND appear to exhibit some dependency upon the presence of AP-1 to enable HIF binding to chromatin. Although importantly, 55% of the HIF-1 β sites that are co-occupied with JUND exhibited no statistically differential binding in response to AFos, and are thus deemed to be unaffected by the ablation of AP-1. It should also be noted that a small number of HIF-1 β sites increased in binding upon the loss of AP-1, however when HIF binding was affected, it was generally a decrease in HIF-1 β binding being that was observed – thus suggesting a cooperative relationship between HIF and AP-1, as opposed to a competitive relationship.

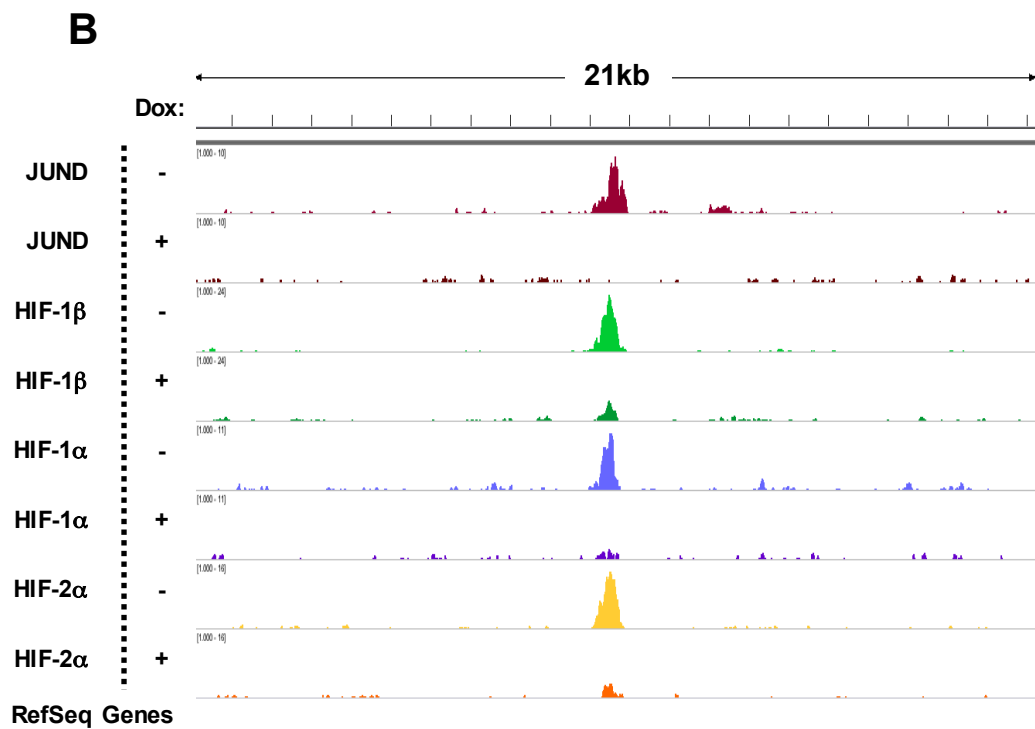
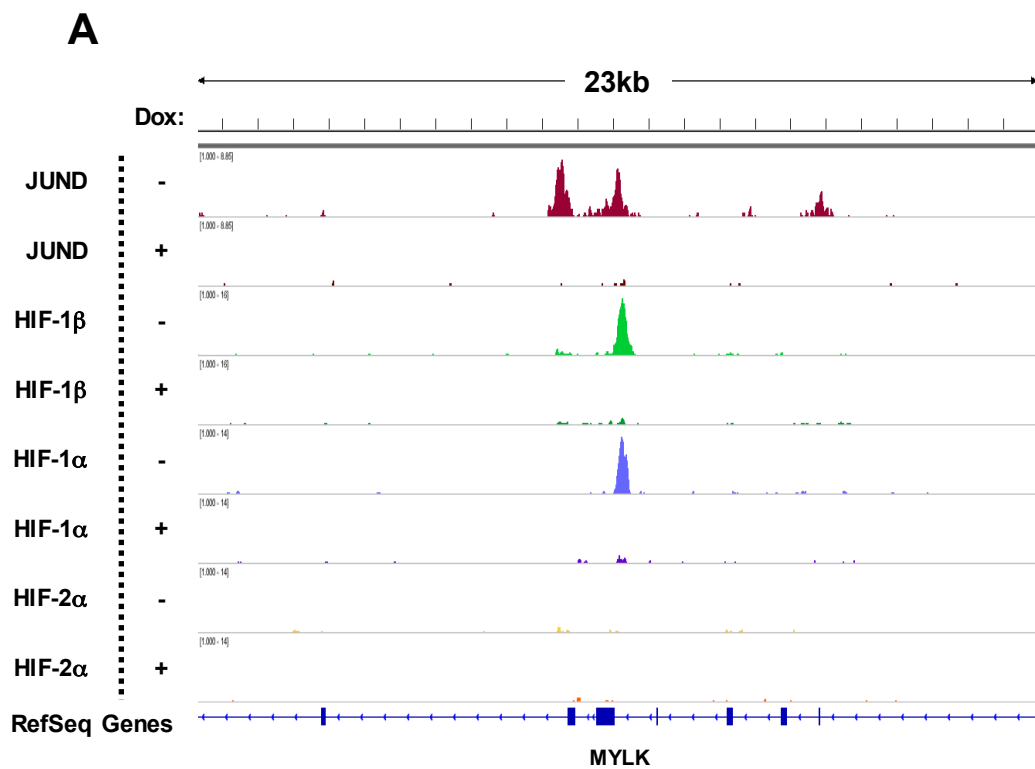


Figure 4. Identification of differentially bound HIF-1 β sites in the presence of AFos.

(A) A 23 kb genomic region within the *MYLK* gene depicting a representative biological replicate JUND (red), HIF-1 β (green), HIF-1 α (blue) and HIF-2 α (orange) ChIP-seq samples in the presence and absence of AFos. This HIF-1 α preferentially bound site evokes a decrease in HIF-1 β and HIF-1 α binding in concordance with the loss of JUND binding in response to AFos induction.

(B) A 21 kb intergenic region depicting a representative biological biological replicate JUND (red), HIF-1 β (green), HIF-1 α (blue) and HIF-2 α (orange) ChIP-seq samples in the presence and absence of AFos. This locus that is bound by all HIF isoforms exhibits a decrease in HIF-1 α , HIF-2 α and HIF-1 β signal intensity in response to AFos induction. Whilst this JUND signal is not called by MACS, there is clear JUND signal in the absence of AFos that is subsequently lost following AFos induction.

5.2.5 Characterisation of the pan-genomic distribution of HIF binding sites with respect to the nearest transcriptional start site

AP-1 has previously been described as binding promoter-distal enhancers in preference to promoter-proximal sites (Zanconato et al., 2015). Therefore, since AP-1 is able to promote HIF-binding, it seems likely that it will do so more often at promoter-distal sites and that it will therefore contribute, at least in part to the distribution of HIF-binding sites around the promoter. I therefore examined the effect of AFos expression on the binding distribution of HIF around the nearest TSS, using HIF-1 β binding as a surrogate for both HIF-1 and HIF-2 binding.

I first confirmed the promoter-distal distribution of AP-1 binding using JUND ChIP-seq analyses performed in vehicle-treated HKC8 cells that were not expressing AFos. 92.3% of JUND sites lay more than 5 kb from the nearest TSS and 49.5% lay more than 50 kb away (Figure 5A). Similarly, HIF-1 β sites that overlap a JUND binding signal showed a comparable promoter-distal distribution to that observed for all JUND sites (Figure 5B-C). However, there was no difference in distribution between HIF/JUND co-occupied sites wherein HIF-1 β binding was or was not affected by AFos.

Since induction of AFos depletes HIF-binding at sites co-bound by both JUND and HIF, the loss of a subset of promoter-distal HIF-binding sites should result in the remaining binding sites occupying a more promoter-proximal distribution. I therefore examined the distribution of HIF-1 β binding sites in the absence and in the presence of AFos. In vehicle-treated cells, HIF-1 β occupied a relatively promoter-distal distribution (Figure 5D). As predicted, ablation of AP-1 binding following the induction of AFos, lead to HIF-1 β occupying a more

promoter-proximal binding distribution, confirming the ability of AP-1 to influence the distribution of HIF-binding sites around the promoter (Figure 5E).

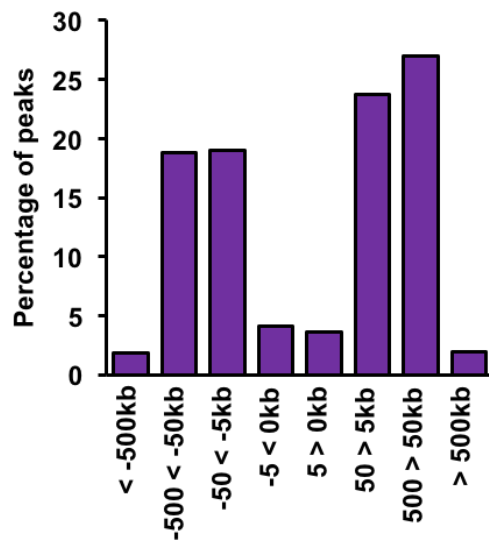
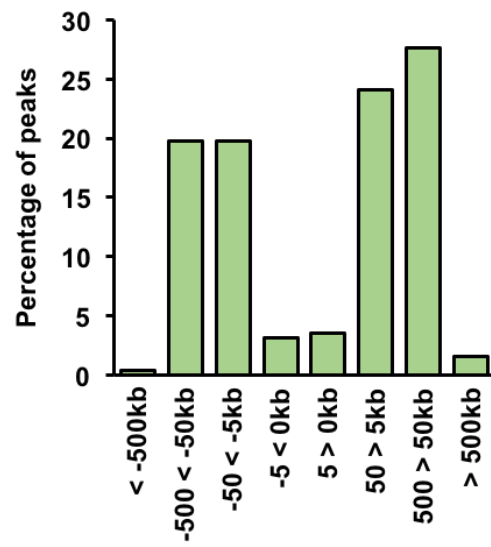
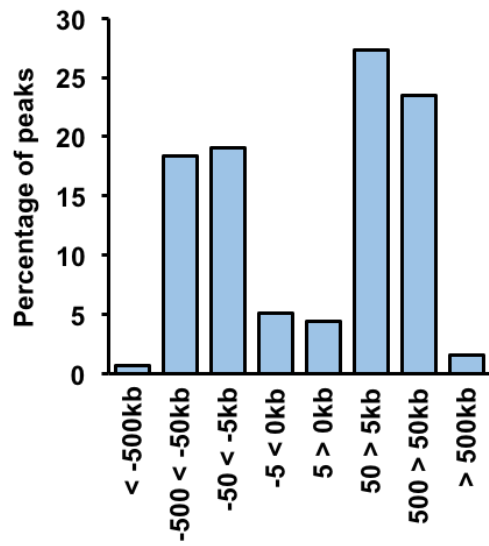
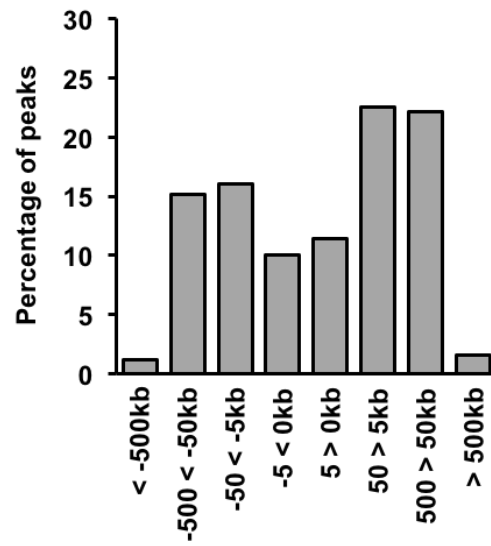
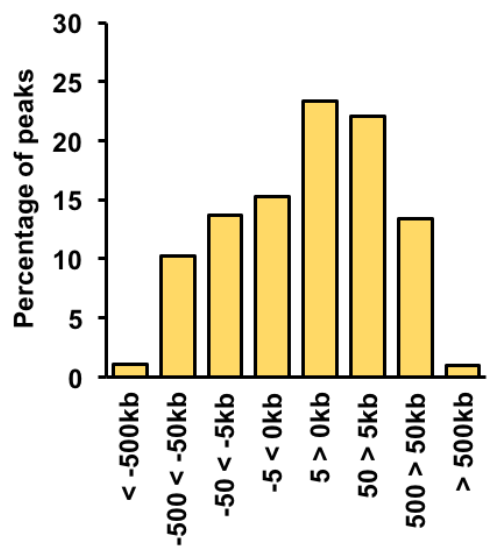
A**B****C****D****E**

Figure 5. Distribution of HIF-1 β and JUND binding sites relative the TSS of the nearest gene.

(A) Histogram depicting the proportion of JUND peak summits falling within defined bins based upon the peak summit's distance from the TSS of the nearest gene. Sites represented here are the totality of JUND sites in vehicle-treated RCC4 pTRIBZ-AFos cells

(B) As in (A), although sites assayed here are HIF-1 β sites that are shared with JUND and exhibiting differential binding of HIF-1 β following AFos induction.

(C) As in (A), although sites assayed here are HIF-1 β sites that are shared with JUND but not exhibiting differential binding of HIF-1 β following AFos induction.

(D) As in (A), although sites assayed here are the totality of HIF-1 β sites in vehicle-treated RCC4 pTRIBZ-AFos cells.

(E) As in (A), although sites assayed here are the totality of HIF-1 β sites in doxycycline-treated RCC4 pTRIBZ-AFos cells.

5.2.6 Attributing biological processes to HIF and AP-1 cooperativity

In the previous section of this chapter, I explored whether the distance between a HIF binding site and the nearest gene could enable prediction of whether HIF and AP-1 may be functioning cooperatively at a site. Here I wished to assess the functionality of the HIF and AP-1 cooperativity by firstly associating the binding sites to Gene Ontology (GO) Biological Process, and then subsequently assessing whether the HIF-1 β and AP-1 co-occupied binding sites were exerting functional regulation of the nearest gene through the use of RNA-seq and gene set enrichment analysis (GSEA). To aid these analyses, I stratified HIF binding sites into three distinct categories as follows:

1. HIF-1 β binding sites in vehicle-treated RCC4 pTRIBZ-AFos cells that are not overlapping a JUND binding site (492 sites)
2. HIF-1 β binding sites in vehicle-treated RCC4 pTRIBZ-AFos cells that overlap a JUND binding site, wherein HIF-1 β binding affinity is not affected by induction of AFos (317 sites)
3. HIF-1 β binding sites in vehicle-treated RCC4 pTRIBZ-AFos cells that overlap a JUND binding site, wherein HIF-1 β exhibits a decrease in binding affinity following induction of AFos (253 sites)

Further examining the three sub-classes of HIF-1 β binding site outlined in the previous section of this chapter, I annotated each binding site by its nearest gene, thus generating three distinct gene lists. GO Biological Process gene list enrichment analysis was performed using ENRICH (Chen et al., 2013). The top five GO Biological Process terms enriched among each of the three tested gene lists are illustrated in table format (Figure 6-8). Enriched terms are ranked

according to their 'Combined Score' that incorporates a *p*-value derived from a Fisher exact test and the deviation from expected rank z-score.

GO Biological Process terms such as those that relate to the 'Response to hypoxia' or 'Response to low oxygen' are enriched among the genes closest to the HIF binding sites whereby AP-1 is not influencing HIF binding (Figures 6,7). These GO term enrichments are to be expected given that the nearest gene to a HIF binding site tends to be induced in response to hypoxia (Schodel et al., 2011).

However, the most enriched GO Biological Process terms among genes proximal to HIF-1 β binding sites that are co-occupied by AP-1, whilst also exhibiting dependency upon AP-1 for binding, are different to those enriched among the other sub-classes of HIF-1 β binding sites and different to what would be expected to be enriched among HIF target genes (Figure 8). The most strongly enriched GO terms are those associated with cell motility, such as 'Induction of positive chemotaxis' (GO:0050930), as well as terms associated with down-regulation of protein phosphorylation, such as 'Negative regulation of phosphorylation' (GO:0042326). The enrichment of these GO terms suggests that the DNA-binding cooperativity of HIF and AP-1 may be responsible for regulating distinct sets of HIF target genes that have not previously been characterised. Therefore, HIF and AP-1 may be functioning to influence specific phenotypic behaviour of a cell, potentially related to cell motility and the negative regulation of protein phosphorylation.

GO Term Name	GO Term ID	Overlap	Combined Score	Genes
Response to decreased oxygen levels	GO:0036293	21/245	18.48	BDNF CITED2 EPAS1 UBE2D2 STC1 PLOD2 NOL3 BACH1 VEGFA TXN2 BMP2 ALKBH5 FAM162A STC2 DDIT4 UBC LONP1 HMOX1 ALDOC P4HB PAM
Response to hypoxia	GO:0001666	21/241	18.46	EPAS1 CITED2 BDNF UBE2D2 STC1 PLOD2 NOL3 BACH1 VEGFA TXN2 BMP2 ALKBH5 FAM162A STC2 DDIT4 UBC LONP1 HMOX1 ALDOC P4HB PAM
Response to oxygen levels	GO:0070482	21/259	17.37	EPAS1 BDNF CITED2 UBE2D2 STC1 PLOD2 NOL3 BACH1 VEGFA TXN2 BMP2 ALKBH5 FAM162A STC2 DDIT4 UBC LONP1 HMOX1 ALDOC P4HB PAM
Glycolytic process	GO:0006096	8/44	11.94	GPI LDHA PFKFB4 PKM ALDOC ENO1 ENO2 HK2
Regulation of transcription from RNA polymerase II promoter in response to hypoxia	GO:0061418	6/29	11.13	CITED2 EPAS1 UBE2D2 UBC BACH1 VEGFA

Figure 6. GO Biological Processes enriched in HIF-1 β sites that are not overlapping JUND binding sites.

The nearest gene to the peak summits of HIF-1 β sites that are not overlapping JUND binding sites were established. The top five GO Biological Processes are depicted in the table and are sorted by 'Combined Score'. Combined score is computed by taking the log of the p -value from the Fisher exact test and multiplying that by the z-score of the deviation from the expected rank.

GO Term Name	GO Term ID	Overlap	Combined Score	Genes
Response to decreased oxygen levels	GO:0036293	15/245	14.77	ARNT2 BDNF CITED2 EPAS1 CAV1 ITPR1 KCNA5 LIMD1 NDRG1 NPEPPS UCN2 KCNMA1 DDIT4 CA9 ANGPTL4
Response to hypoxia	GO:0001666	15/241	14.75	ARNT2 EPAS1 CITED2 BDNF CAV1 ITPR1 KCNA5 LIMD1 NDRG1 NPEPPS UCN2 KCNMA1 DDIT4 CA9 ANGPTL4
Response to oxygen levels	GO:0070482	15/259	13.93	ARNT2 EPAS1 BDNF CITED2 CAV1 ITPR1 KCNA5 LIMD1 NDRG1 NPEPPS UCN2 KCNMA1 DDIT4 CA9 ANGPTL4
Angiogenesis	GO:0001525	15/236	13.76	VAV3 NRP1 EPAS1 CAV1 ITGB3 FN1 TGFA THY1 CYR61 RHOB PTPRB MYH9 ANGPTL4 TGFB1 RAPGEF3
Negative regulation of peptidyl-threonine phosphorylation	GO:0010801	4/11	13.12	DMTN DDIT4 CALM1 SMAD7

Figure 7. GO Biological Processes enriched in HIF-1 β sites that are shared with JUND but not exhibiting differential binding of HIF-1 β following AFos induction.

The nearest gene to the peak summits of HIF-1 β sites that are shared with JUND, but not exhibiting differential binding of HIF-1 β following AFos induction, were established. The top five GO Biological Processes are depicted in the table and are sorted by 'Combined Score'. Combined score is computed by taking the log of the p -value from the Fisher exact test and multiplying that by the z-score of the deviation from the expected rank.

GO Term Name	GO Term ID	Overlap	Combined Score	Genes
Induction of positive chemotaxis	GO:0050930	4/15	10.04	ARTN NTF3 SCG2 VEGFA
Negative regulation of phosphorylation	GO:0042326	15/325	9.56	JUN KIRREL GADD45B PRKAG2 IRS2 PTPRJ DYNLL1 GFRA2 NLRP12 DUSP10 LATS2 CAMK2N1 MYC NTF3 TRIB1
Positive regulation of locomotion	GO:0040017	13/304	8.97	ITGB1 ITGA3 SERPINE1 IRS2 ETS1 EGFR PTK2 VEGFA ARTN NTF3 ROR2 DRD1 SCG2
Negative regulation of protein phosphorylation	GO:0001933	12/260	8.88	JUN NLRP12 KIRREL DUSP10 LATS2 GADD45B CAMK2N1 NTF3 PRKAG2 PTPRJ TRIB1 GFRA2
Positive regulation of positive chemotaxis	GO:0050927	4/26	8.73	ARTN NTF3 SCG2 VEGFA

Figure 8. GO Biological Processes enriched in HIF-1 β sites that are shared with JUND and exhibit differential binding of HIF-1 β following AFos induction.

The nearest gene to the peak summits of HIF-1 β sites that are shared with JUND, and exhibit differential binding of HIF-1 β following AFos induction, were established. The top five GO Biological Processes are depicted in the table and are sorted by 'Combined Score'. Combined score is computed by taking the log of the *p*-value from the Fisher exact test and multiplying that by the z-score of the deviation from the expected rank.

5.2.7 Association of HIF and AP-1 binding sites with gene expression

I have previously identified a subset of HIF binding sites in which HIF exhibits a dependence upon AP-1 for binding to DNA. I subsequently demonstrated that the nearest genes to these sites enrich among GO Biological Process terms relating to motility and chemotaxis. I therefore wanted to establish whether the apparent cooperative relationship between HIF and AP-1 is involved in the control of gene expression. RNA-seq was performed in the presence and absence of the AP-1 dominant-negative, and gene set enrichment analysis (GSEA) was subsequently performed in order to assess the determine whether the genes nearest the sites wherein HIF binding is suppressed in response to AFos are among the genes down-regulated in response to AFos.

Total RNA was harvested from RCC4 pTRIBZ-AFos cells that had been cultured in normoxia, whilst being treated with vehicle or doxycycline (1 μ g/ml) for 48 hours. Next generation sequencing libraries were prepared from RNA by polyA⁺ selection, and RNA-seq was subsequently performed. RNA was harvested in biological triplicate for each condition. DESeq2 analysis was performed to determine the vehicle- vs. doxycycline-treated fold change across all detected transcripts, in addition to differential expression significance. Genes were ranked by a value combining differential expression significance according to DESeq2 and the fold change (Xiao et al., 2014). The binding sites within the three previously described categories of HIF binding site were annotated by their nearest protein coding gene, and these genes were subsequently assessed for enrichment among the ranked RNA-seq gene list. With respect to an adjusted *p*-value of threshold of 0.05, 3866 genes were significantly up-regulated following a

48-hour induction of AFos, whereas 3923 genes were significantly down-regulated.

The right panels of Figure 9 (B, D, F) illustrate associations between the different classes of HIF binding site and responses to suppression of HIF-1 α and HIF-2 α by siRNA (Unpublished microarray data generated by former Ratcliffe Group member, Raju Raval). The left panels of Figure 9 (A, C, E) illustrate associations between the different classes of HIF binding gene and responses to ablation of AP-1 by AFos.

Neither the HIF-1 β sites that are not co-occupied by JUND ($p = 0.37$), nor those that are co-occupied by JUND but not exhibiting any dependence on the presence of AP-1 for DNA-binding ($p = 0.74$), enriched among genes negatively regulated during the induction of AFos (Figure 9A and 9C). However, these sites did enrich among genes negatively regulated in RCC4 cells that were transfected with siRNA against HIF-1 α and HIF-2 α (Both $p < 0.001$; Figure 9B and 9D), therefore demonstrating that they are functional HIF binding sites, but that their function is independent of AP-1.

The HIF-1 β binding sites that exhibit a dependence upon AP-1 for DNA binding enrich among genes negatively regulated following the induction of AFos ($p < 0.001$; Figure 9E). This suggests that AP-1 binding at these HIF sites contributes to the positive regulation of the nearby genes. These HIF-1 β binding sites also enrich among genes negatively regulated in RCC4 cells that are transfected with siRNA against HIF-1 α and HIF-2 α ($p = 0.03$; Figure 9F).

During the previous section in this chapter, I showed that the HIF-1 β binding sites that exhibit a dependence upon AP-1 for DNA binding, neighbour genes that are enriched within GO Biological Process terms related to

chemotaxis and locomotion. Given the relation of these phenotypes to cancer (Hanahan and Weinberg, 2000), I assessed whether the enriched genes are differentially regulated in the presence vs. the absence of AFos. Of the 13 enriched genes, 8 eight were significantly down-regulated following a 48-hour treatment of cells with doxycycline to induce AFos (*NTF3*, *SCG2*, *ITGB1*, *ITGA1*, *SERPINE1*, *ETS1*, *PTK2*, *ROR2*). 2 genes were unchanged (*ARTN*, *DRD1*), whereas 3 genes significantly increased (*VEGFA*, *IRS2*, *EGFR*).

Taken together, these data corroborate published data in that the nearest gene to a HIF binding site tends to be positively regulated by HIF. Moreover, AP-1 only influences the expression of HIF target genes in situations where AP-1 and HIF are co-occupying a binding site, and wherein the DNA binding of AP-1 is contributing to the capacity of HIF to bind DNA.

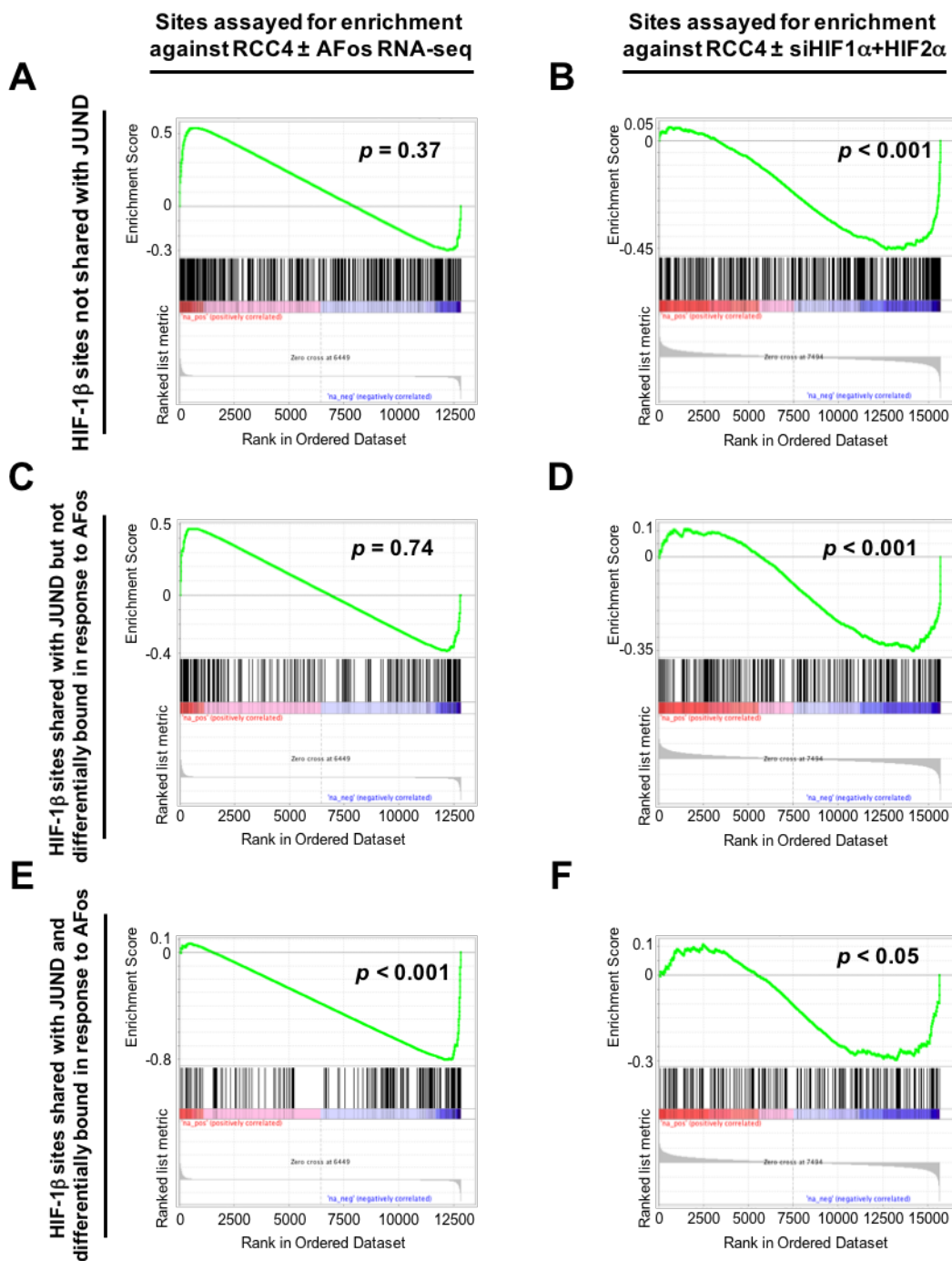


Figure 9. Functional association of JUND and HIF-1 β sites in RCC4 pTRIBZ-AFos cells.

(A) Association of the HIF-1 β sites that are not overlapping JUND binding sites in vehicle-treated RCC4 pTRIBZ-AFos cells, with genes regulated in response to 48 hours doxycycline treatment to induce AFos, via gene set enrichment analysis (GSEA) ($p = 0.37$).

(B) Association of the HIF-1 β sites that are not overlapping JUND binding sites in vehicle-treated RCC4 pTRIBZ-AFos cells, with genes regulated by siRNA-mediated knockdown of HIF-1 α and HIF-2 α , via gene set enrichment analysis (GSEA) ($p < 0.001$).

(C) Association of the HIF-1 β sites that are overlapping JUND binding sites but do not exhibit differential binding of HIF-1 β following AFos induction, in vehicle-treated RCC4 pTRIBZ-AFos cells, with genes regulated in response to 48 hours doxycycline treatment to induce AFos, via gene set enrichment analysis (GSEA) ($p = 0.74$).

(D) Association of the HIF-1 β sites that are overlapping JUND binding sites but do not exhibit differential binding of HIF-1 β following AFos induction, in vehicle-treated RCC4 pTRIBZ-AFos cells, with genes regulated by siRNA-mediated knockdown of HIF-1 α and HIF-2 α , via gene set enrichment analysis (GSEA) ($p < 0.001$).

(E) Association of the HIF-1 β sites that are overlapping JUND binding sites and exhibit differential binding of HIF-1 β following AFos induction, in vehicle-treated RCC4 pTRIBZ-AFos cells, with genes regulated in response to 48 hours doxycycline treatment to induce AFos, via gene set enrichment analysis (GSEA) ($p < 0.001$).

(F) Association of the HIF-1 β sites that are overlapping JUND binding sites and exhibit differential binding of HIF-1 β following AFos induction, in vehicle-treated RCC4 pTRIBZ-AFos cells, with genes regulated by siRNA-mediated knockdown of HIF-1 α and HIF-2 α , via gene set enrichment analysis (GSEA) ($p < 0.05$).

5.3 Discussion

This chapter describes novel insights into the pan-genomic functional relationship between HIF and the basic leucine zipper transcription factor, AP-1. Specifically, I began by presenting data that illustrates the levels of binding site coincidence exhibited between the two transcription factors. Subsequently, through the utilisation of a pan-genomic intervention against AP-1 (Olive et al., 1997), I illustrated to what extent HIF is dependent upon AP-1 for chromatin binding at co-occupied loci. I subsequently assessed how this DNA binding cooperativity translates into effects on gene expression.

5.3.1 Ablation of AP-1 DNA binding perturbs the pan-genomic binding of HIF

My findings reveal that JUND preferentially coincides with HIF-2 sites relative to HIF-1. This finding corroborates the preferential AP-1 motif enrichment within HIF-2 sites in RCC4 cells that was described in Chapter 3. However, in spite of the greater association of AP-1 with HIF-2 relative to HIF-1, intervention against AP-1 using the dominant-negative, AFos, revealed that subsets of both HIF-1 and HIF-2 binding sites could be affected by AP-1 perturbation.

Induction of AFos in RCC4 cells led to a reduction in HIF binding at a significant subset of sites. The majority of these sites coincided with local AP-1 binding, thus suggesting that AP-1 is exerting an effect on HIF recruitment to chromatin in *cis*. Prior studies have detailed the local cooperative relationship between HIF and AP-1 in the regulation of *VEGFA* and *EDN1* (Damert et al., 1997; Yamashita et al., 2001). Later studies predicted genome wide cooperativity of AP-1 and HIF following the observation of enriched AP-1 motifs within HIF binding sites genome wide (Villar et al., 2012; Salama et al., 2015). However,

until now no study has demonstrated widespread binding cooperativity of HIF and AP-1.

Importantly, AP-1 perturbation is not affecting all HIF binding sites. This suggests that AP-1 is not influencing transcription factor binding across all *cis*-regulatory elements, rather there are specific effects of AP-1 on HIF at selected sites. The majority of HIF binding events that are affected by AP-1 perturbation are sites whereby there is proximal binding of AP-1. Although interestingly, HIF binding is not affected by AP-1 at all sites in which AP-1 is bound proximally.

A *cis* effect of AP-1 on HIF binding may arise be a result of AP-1 acting as a pioneer factor, as has been reported in the case of transcriptional cooperativity between AP-1 and the glucocorticoid receptor (GR) (Biddie et al., 2011). Pan-genomic binding of the GR was determined via ChIP-seq; effects on gene expression were determined via microarray, and genome-wide chromatin accessibility changes were determined via DNaseI-seq. Interestingly, AP-1 was shown to maintain chromatin accessibility at GR-target loci, thus facilitating GR recruitment and GR target gene transactivation. Given that the vast majority of HIF sites that are influenced by AP-1 presence are proximal to an AP-1 binding site, it is possible that AP-1 is performing a similar pioneer transcription factor role for HIF as it is in the case of the GR. Future experiments should entail the use of DNaseI-, FAIRE- or ATAC-seq in order to assay genome-wide changes in chromatin accessibility in response to AFos in RCC4 cells, and subsequently assess whether changes in accessibility are coinciding with the suppression of HIF binding. Moreover, this potential mechanism may relate to the recent findings by the Ratcliffe group in that HIF binds to pre-formed chromatin loops (Platt et al., 2016). Functioning as a pioneer factor, AP-1 may be at least partially responsible

for maintaining these loops, and thus maintaining a functional interaction between HIF-bound enhancers and HIF target genes.

AP-1 components have previously been shown to physically interact with HIF-1 α (Alfranca et al., 2002; Yu et al., 2009), therefore a potential mechanism underpinning the dependence of HIF on AP-1 for binding to a subset of its sites could be via a HIF/AP-1 protein-protein interaction. The absence of AP-1 at a subset of HIF target loci could severely reduce the affinity of HIF for binding to these sites.

Notably, there was a small subset of HIF binding sites whereby the genome-wide ablation of AP-1 binding reduced HIF binding in spite of AP-1 not being bound proximally. A caveat to this statement was discussed earlier in the Results section of this chapter in that MACS under-called JUND binding sites. However, there was not significant JUND ChIP-seq signal at all of these sites. This therefore indicates that AP-1 may also influence AP-1 chromatin binding in *trans*. Reasons for this *trans* effect on HIF binding could be that AP-1 down-regulation evokes a down-regulation of a transcription factor that is required for maximal HIF binding at a subset of sites. Or, AP-1 could be influencing the capacity for HIF to bind via long distance chromatin-chromatin interactions. A caveat to these postulations is that this represents only a small proportion of the HIF binding sites that are affected by AP-1 intervention. Also, I have used a surrogate of JUND for genome-wide AP-1 binding, therefore it is possible that alternate AP-1 components, independent of JUND, are comprised within the AP-1 dimers bound at these sites.

5.3.2 Cooperative binding of HIF and AP-1, and its potential relation to cancer

The nearest genes to the HIF binding sites whereby AP-1 was bound locally, and HIF binding was reduced following AP-1 intervention, showed a remarkable association with GO terms associated with cell motility, chemotaxis and locomotion. This is unlike the remaining HIF binding sites that enrich for GO terms associated with expected biological processes such as the 'response to low oxygen'. There has never been a report that has previously defined a subset of HIF binding sites associated with this phenotype. A significant proportion of the genes comprised within these cell motility-related GO terms are also down-regulated in response to AP-1 intervention, and separately, siRNA-mediated depletion of HIF-1 α and HIF-2 α . This suggests that HIF and AP-1 may be regulating these genes in a synergistic fashion.

Hypoxia has been linked to cancer cell motility and invasion in a number of contexts, including breast cancer (Muñoz-Nájjar et al., 2006; Gilkes et al., 2014) and ovarian cancer (Shi et al., 2008). Likewise, the highly pleiotropic transcription factor, AP-1, has been demonstrated to promote tumour cell motility in a vast array of contexts (Reviewed here: Ozanne et al., 2007). However, these factors have not previously been demonstrated to act in concert with one another to exert effects on motility and cancer cell invasion. Future assays should be performed in order to understand whether HIF and AP-1 act synergistically through a defined gene set to influence this cancer phenotype, and whether this potential pro-tumourigenic behaviour is specifically associated with clear cell renal carcinoma. Should this indeed be a HIF/AP-1 influenced phenotype,

therapeutic small molecule antagonists could feasibly be developed to inhibit the functional interaction between HIF and AP-1 in renal cancer.

Chapter 6: Final Discussion

Transcriptional responses orchestrated by HIF are wide-ranging and involve the direct transactivation of many hundreds of genes that are central to cellular adaptation to hypoxia. However, despite the large numbers of documented HIF binding sites on chromatin, only a small proportion of the several million HIF binding motifs (HREs) in the genome actually bind HIF. This is true, even when accessibility as evidenced by DNA-seq, ATAC-seq or FAIRE-seq is taken into account and indicates that other factors, beyond consensus motif and DNA-accessibility, help govern the selection of HIF-binding sites. This raises important questions as to which binding sites are selected, how these sites differ in different settings, whether different HIF isoforms behave differently and what mechanisms underlie binding site selectivity. These questions are of fundamental importance not only to understanding the physiological response to hypoxia, but also to understanding transcriptional selectivity more generally. They also have important implications for understanding how the HIF pathway is shaped (and could potentially be therapeutically remodelled) in diseases such as cancer in which HIF activation may have diverse effects. This thesis, has addressed these questions, using recent advances in high-throughput next-generation sequencing technologies coupled to chromatin immunoprecipitation. Patterns of HIF binding across the genome and their association with potential effectors of binding site selection were then analysed using novel computational and bioinformatic approaches.

6.1 Canonical vs. non-canonical binding of HIF

One question arising out of prior work is whether HIF- α is able to bind chromatin without HIF-1 β , or vice versa. Indeed, previous studies have suggested that

endogenous HIF- α , binding may bind to chromatin at specific sites without HIF-1 β (Gustafsson et al., 2005; Xiang et al., 2015). Furthermore, in overexpression studies, a more extensive distribution of HIF- α binding (particularly HIF-2 α binding) without HIF-1 β was observed (Salama et al., 2015). However, other work (Schodel et al., 2011) had suggested that in MCF-7 cells endogenous HIF- α largely bound chromatin in association with HIF-1 β .

I therefore extended the analysis of endogenous HIF- α and HIF-1 β binding to three additional cell lines. A primary observation was that simply defining HIF-binding sites by applying the commonly used ChIP-seq peak caller, MACS, has a strong tendency to over-estimate differences in binding sites between two datasets. Specifically, using highly stringent statistical thresholds to define high-confidence binding site occupancy led to the exclusion of many HIF-1 β sites with substantial levels of quantitative ChIP-seq signal that correlated strongly with HIF- α signal. When more quantitative methods of analysis were applied to these sites, there was evidence of HIF-1 β binding at almost every HIF-1 α or HIF-2 α site assayed, and vice versa, in each of the three cell lines examined. Further quantitative analysis of HIF- α and HIF-1 β signal in the three cell lines revealed a strong correlation between total HIF- α signal and HIF-1 β signal with only minor deviations from a 1:1 ratio. Taken together this indicates that at least in these settings, endogenous HIF- α always binds chromatin in association with HIF-1 β and vice versa.

These findings are in contradistinction to a re-analysis of the published overexpression studies (Salama et al., 2015), using the same methodologies, which confirmed the findings of the previous work demonstrating that

overexpressed HIF- α is able to bind without HIF-1 β . It is possible that this observation is simply an artefact of very high expression levels, which force HIF- α to bind alone. However, it is equally possible that endogenous levels of HIF- α would be able to bind chromatin on their own, but they are inhibited from doing so because endogenous HIF- α only exists in complex with HIF-1 β . Overexpression of HIF- α to levels in excess of HIF-1 β might then reveal this latent property. This hypothesis could be tested by suppressing HIF-1 β to levels below those of HIF- α using RNAi or deleting the gene using CRISPR-Cas9. However, this potential mechanism could not explain the inability of endogenous HIF-1 β to bind without HIF- α , since endogenous HIF-1 β is present in excess of endogenous HIF- α , yet is still unable to bind without HIF- α .

Another caveat is that the HIF-1 β antibody may not be entirely specific for HIF-1 β alone. Several other HIF-1 β (ARNT) homologs, ARNT2, ARNTL1 and ARNTL2 are also present within the human genome and could potentially dimerise with HIF- α even though the transcripts of these genes are generally expressed at a lower level than those for HIF-1 β . If the HIF-1 β antibody also recognised these homologous proteins then they would not be distinguished separately in my ChIP-seq analyses. However, if this was the case, then again I would expect to observe residual HIF- α binding if the HIF-1 β gene was deleted by CRISPR-Cas9.

It is also possible that HIF-1 β might dimerise with other partners than HIF, but that these are not present in the cell lines tested under the tissue culture conditions used. For example, it is well recognised that AhR is able to bind chromatin in association with HIF-1 β . However, AhR mRNA levels were

considerably lower than those for either HIF-1 α or HIF-2 α and the cells were not cultured in the presence of an appropriate ligand to stabilise AhR protein. Again the ability of HIF-1 β to bind without HIF- α could be tested by engineering combined deletion of both HIF- α isoforms simultaneously using CRISPR-Cas9.

Finally, it is also possible that either HIF- α or HIF-1 β could bind chromatin in association with a dimerisation partner that obscures the epitopes recognised by the antibody used in the ChIP-seq analyses. This is unlikely, since bHLH-PAS transcription factors usually dimerise through their PAS domains, which would have been obscured in the HIF- α /HIF-1 β heterodimer. Furthermore, the antibodies used were all polyclonal antibodies generated to large C-terminal regions of each protein well away from their dimerisation domains.

6.2 Cell type specificity of HIF binding sites and binding site distributions

Although the HIF transcription factor is ubiquitous across all nucleated cells within the body, the transcriptional response to hypoxia is highly variable (Ortiz-Barahona et al., 2010). The extent to which HIF binding sites are conserved across cell lines, or specific to one cell line, has been investigated previously (Villar et al., 2012). However, these datasets were predominantly obtained from ChIP-on-chip assays. I therefore sought to readdress the question of HIF binding site conservation using high depth ChIP-seq, which allows the interrogation of all sites across the genome irrespective of distance from annotated promoters, whilst comparing and contrasting the behavior of the two major HIF isoforms, HIF-1 and HIF-2.

Corroborating the previous published data, my findings show that the majority of HIF binding sites are cell specific. Cell type specificity was greater for HIF-2 than HIF-1, with only 3.1% of HIF-1 and 1.1% of HIF-2 being 'called' by the MACS peak-calling algorithm across all of the four cell lines studies in this work. This was despite all four cell lines expressing both sub-units. Studies to date have emphasised the cell type specificity of expression of the two isoforms of HIF- α . HIF-1 α is expressed in the large majority of cell and tissue types whereas HIF-2 has a much more restricted expression pattern. Interestingly, in cancer HIF-2 is often de-repressed, as occurs during the development of clear cell renal carcinoma from the renal tubular epithelium. The current work indicates that despite an apparently identical core consensus binding sequence, a further level of tissue or cell specificity in the HIF response is determined at the level of HIF binding and that again, this is somewhat greater for HIF-2 than HIF-1.

One of the most striking findings in this work was the finding that HIF-1 and HIF-2 binding sites were differently distributed with respect to gene promoters. This was evident when simple physical distances from transcriptional start sites were measured, and also when co-incidence with histone 'marks' that are typical of enhancers and promoters were measured. Whilst HIF-1 was strikingly enriched at promoters and showed a strong association with the 'promoter mark' H3K4me3, HIF-2 was distributed more widely at distances remote from promoters and coincided clearly with the 'enhancer mark' H3K4me1.

This difference in the pan-genomic distribution of binding was observed in each cell type; in each case HIF-1 was more promoter-proximal than HIF-2, and despite the binding sites themselves being largely different across cell types, the HIF-isoform-specific distributions remained similar. Even when HIF-1 and HIF-2

sites were categorised according to the number of cells that the site is conserved within, HIF-1 and HIF-2 binding sites in each category (of cell-specificity) still exhibit these contrasting distributions with respect to their propensity to bind close to promoters or at promoter-distal loci.

Interestingly neither distribution conforms to the distribution of core HREs in open chromatin, suggesting that in each case factors other than this sequence must operate actively to promote DNA binding. The maintenance of the contrasting promoter-proximal and promoter-distal binding site distributions for HIF-1 and HIF-2 across cell lines, in spite of the fact that only a relatively small proportion of HIF-1 and HIF-2 binding sites are conserved across cell lines, is remarkable and suggests that mechanisms underlying cell-type specificity and enhancer-promoter distribution of binding must be, at least in part, distinct. More specifically, the data suggests the existence of factors that are common across cell types that act to distribute HIF-1 and HIF-2 to their respective promoter-proximal and promoter-distal sites. What these might be is unclear. Neither the motif analysis, nor the associations with transcription factor binding and histone marks described in Chapter 3, has the potential to go beyond defining associations. In future work, consideration should be given to careful immunoprecipitation experiments, which might identify specific interaction partners of HIF-1 and HIF-2 that mediate this contrasting behavior. To my knowledge this type of differential binding distribution amongst closely related transcription factors has not been studied in depth in other systems. Analysis of data deposited by the encode consortium reveals that although some transcription factor families display concordant behavior, contrasting promoter-proximal or promoter-distal biases are also observed among members of other

transcription factor families – the basic leucine zipper (bZIP) transcription factor family being a prime example (Figure 1).

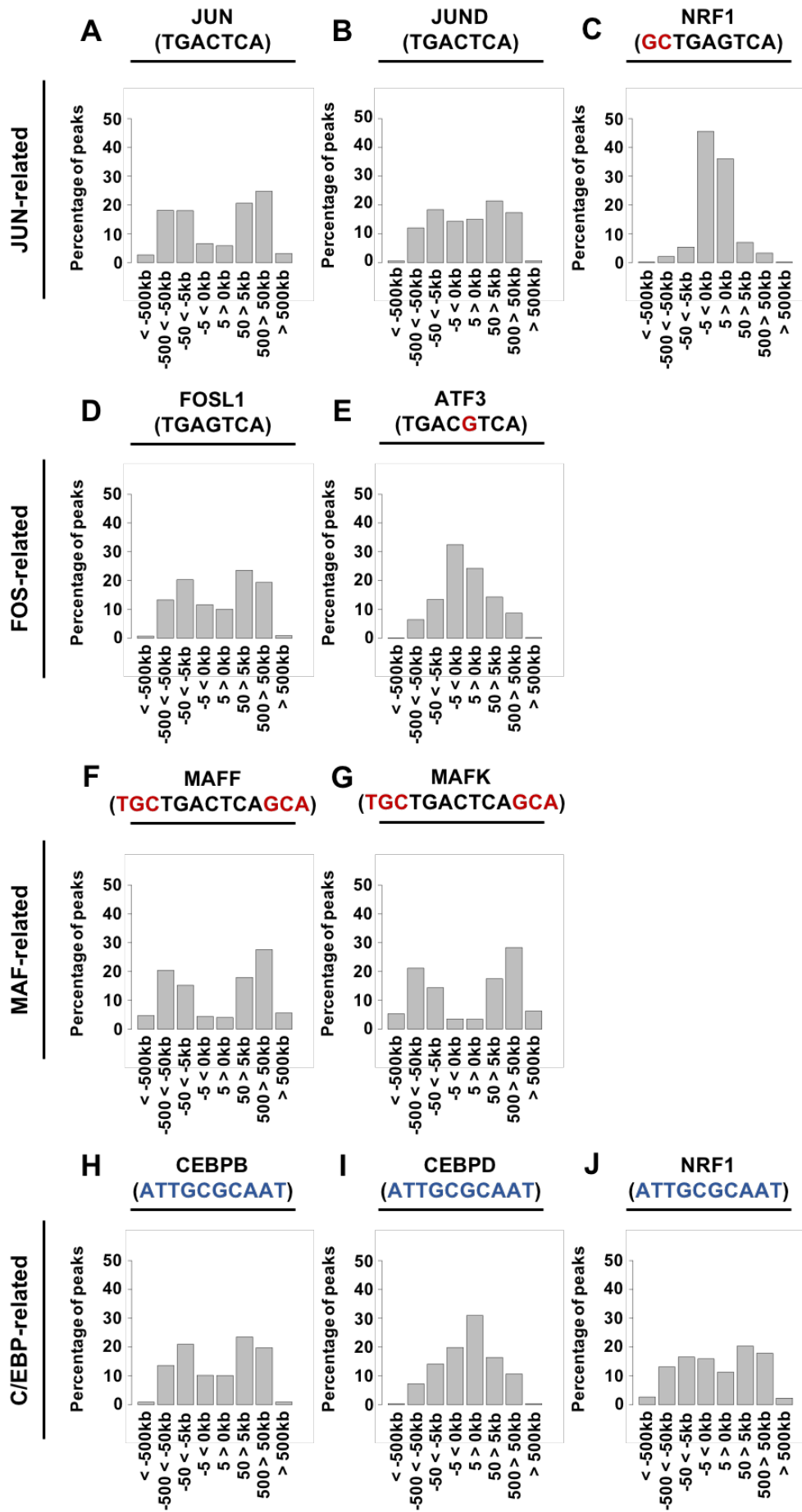


Figure 1. Distribution of bZIP transcription factor binding sites relative to the nearest TSS.

(A) Frequency distribution histogram depicting the proportion of ENCODE-generated JUN binding sites in HepG2 cells, cultured in normoxia, based upon the distance between a site and the TSS of the nearest gene. The JUN consensus motif is shown in brackets.

(B) Histogram as in (A), however ENCODE-generated JUND sites plotted.

(C) Histogram as in (A), however ENCODE-generated NRF1 sites plotted.

(D) Histogram as in (A), however ENCODE-generated FOSL1 sites plotted.

(E) Histogram as in (A), however ENCODE-generated ATF3 sites plotted.

(F) Histogram as in (A), however ENCODE-generated MAFF sites plotted.

(G) Histogram as in (A), however ENCODE-generated MAFK sites plotted.

(H) Histogram as in (A), however ENCODE-generated CEBPB sites plotted.

(I) Histogram as in (A), however ENCODE-generated CEBPD sites plotted.

(J) Histogram as in (A), however ENCODE-generated NRF1 sites plotted.

6.3 Inactivation of HIF- α and the determination of HIF compensatory capacity

Previous studies have demonstrated that increased HIF-1 and HIF-2 levels are associated with increased incidence of tumour metastasis, increased tumour angiogenesis, resistance to tumour therapies, and subsequent poor patient prognosis. A recent study has identified small molecule binding pockets in HIF-1 α and HIF-2 α which may enable selective targeting of the two HIF isoforms for cancer therapy (Wu et al., 2015). A selective HIF-2 α antagonist has been demonstrated as efficacious in abrogating clear cell renal carcinoma growth in pre-clinical models (Wallace et al., 2016; Chen et al., 2016; Cho et al., 2016). The pursuit of selective HIF- α antagonists highlights the need for an increased understanding of the pan-genomic transcriptional activities of HIF-1 and HIF-2 following the inactivation of either HIF-1 α or HIF-2 α . I therefore utilised CRISPR-Cas9 to generate renal proximal tubule cell lines deficient in either HIF-1 α , HIF-2 α or both, and subsequently performed ChIP-seq targeting the present HIF- α isoform and HIF-1 β in -1 α KO and -2 α KO cells to enable interrogation of the HIF binding distributions in these settings.

Remarkably little compensatory binding activities of HIF-1 α or HIF-2 α was observed. Any redistribution of binding sites, or differential changes in ChIP-seq binding signal, were restricted to sites wherein both HIF-1 and HIF-2 are capable of binding in the wild type setting. This was largely consistent with my hypothesis given that in 786O cells, wherein full-length HIF-1 α is absent, HIF-2 α still continues to adopt a promoter-distal binding site distribution (Salama et al., 2015).

A notable caveat to the observations discussed previously arises from the CRISPR-Cas9 HIF- α deficient cell lines being derived from a single cell among a heterogeneous population of aneuploid cells. This means that any described alteration in HIF binding distribution in the knockout cells relative to wild type cells, or lack thereof, could merely be a property of the single cell from which the knockout line is derived. Had profound alterations in binding been observed, further work would need to be directed toward repeating these experiments using a second clone for each knockout line, however repetition is likely unjustified given the largely negative result. Nonetheless, these findings appear to imply that there are distinct mechanisms that apply rigidity to the binding site selection of HIF-1 and HIF-2, respectively. This has significant implications for the continuing development of HIF-1 α and HIF-2 α antagonists, in that isoform-selective therapies would in theory deliver selective hypoxic response effects.

Coupling binding data to transcript profiling by RNA-seq in HIF- α deficient cells has provided new insight into which HIF isoform is absolutely required for hypoxia-induced expression of the gene nearest a *cis*-element capable of binding both HIF isoforms. Interestingly, distance between the shared site and the nearest gene, in addition to the binding intensity of the HIF- α isoform at the shared site, both appear to be predictors of whether the nearby gene is dependent on HIF-1 or HIF-2 for its hypoxic induction. The closer the shared binding site is to the gene, the more likely HIF-1 α is to be bound with greater strength, and the more likely the gene is to be dependent upon HIF-1 for its hypoxic induction. Conversely, the greater the distance between the shared site and the gene, the more likely HIF-2 α is to be bound with greater strength, and the more likely the gene is to be dependent upon HIF-2 for its hypoxic induction.

These findings imply that there are post-binding mechanisms that are distinctly associated with HIF-1 and HIF-2 mediated gene target transactivation, wherein the machinery required for HIF-1 and HIF-2 target gene transactivation may be preferentially localised proximal or distal to the target gene. It is possible that high affinity binding of HIF-1 and HIF-2 is required to coordinate their respective gene transactivation machinery, and again, it is possible that the factors required to recruit HIF-1 and HIF-2 at high affinity are localised proximal or distal to the target gene. Further work exploring the transcription factors bound locally to HIF-1 and HIF-2, in addition to work interrogating the direct interacting partners of HIF-1 and HIF-2, may enable greater understanding of the post-binding mechanisms associated with HIF-1 and HIF-2 gene target transactivation.

6.4 Analysis of transcription factor binding in the vicinity of HIF binding sites

As discussed previously, HIF-1 α and HIF-2 α share significant sequence homology, the same dimerisation partner (HIF-1 β), and both bind the same minimal consensus motif (HRE). Yet, in spite of these similarities, HIF-1 and HIF-2 bind both shared and distinct loci across the genome. To gain greater understanding of what may be driving the selectivity of HIF-1 and HIF-2, sites preferentially bound by either HIF-1 or HIF-2 were analysed for transcription factor binding motifs and ENCODE transcription factor binding site coincidence. HIF-1 and HIF-2 interaction with distinct transcription factors may be a significant contributor to their respective determinants of binding site selection.

Distinct sets of transcription factor motif were found to be enriched within HIF-1 and HIF-2 binding sites, respectively. Interestingly, the enriched motifs

were largely conserved across cell types, thus suggesting that the same transcription factors bind in the local vicinity to HIF-1 and HIF-2 sites regardless of the cell type. Many of the motifs that were enriched within HIF binding sites in HepG2 cells were corroborated by the coincidence of actual binding sites when overlapping ENCODE transcription factor binding data with my HIF ChIP-seq data.

These data suggest that one or more transcription factors, distinctly associated with HIF-1 or HIF-2, are likely to be influencing HIF binding. The motifs observed nearby may also represent transcription factors that HIF is enabling to bind to chromatin.

It is important to note that any motif observation may be observed by chance in that certain nucleotides are preferentially enriched within certain regions of the genome, as exemplified by the fact promoter-proximal regions tend to be GC-rich (Saxonov et al., 2006). A further caveat that needs to remain under consideration when interpreting these data is that the ENCODE ChIP-seq data used to associate with my HIF data was all obtained in normoxia. It is therefore possible that hypoxia influences the expression and, or, activation of a number these factors, therefore their pan-genomic binding sites may not remain the same in normoxia. Nonetheless, such extensive associations between HIF and other transcription factors has not been made previously. Future work should aim to determine the legitimacy of these associations, either by chromatin immunoprecipitation of select factors in hypoxia or VHL-defective cell lines, or by determining whether HIF-1 or HIF-2 interact physically with the selected factors through the use of pull-down assays.

6.5 The effects of AP-1 intervention on HIF binding and function

In Chapter 3, I demonstrated that the AP-1 motif is consistently enriched within HIF-2 binding sites, with a particularly strong enrichment within renal cells. HIF-2 binding sites in HepG2 cells also exhibited a strong coincidence in binding with AP-1 components when associating HIF ChIP-seq data with ENCODE ChIP-seq data. Furthermore, numerous published reports have functionally associated HIF with AP-1 (Laderoute, 2005; Villar et al., 2012; Salama et al., 2015). Given the multiple levels of association between AP-1 and HIF, I introduced an AP-1 dominant-negative (AFos) into RCC4 cells in order to ablate AP-1 binding and subsequently assess the pan-genomic binding of the HIF isoforms in the absence of AP-1.

A major reduction in HIF binding was observed following the induction of AFos, however unexpectedly, the effect of the AP-1 dominant-negative was not restricted to HIF-2 α - HIF-1 α was also affected. Whilst the magnitude of the effect was different at different HIF binding sites, the suppression of HIF binding was predominantly associated with proximal AP-1 binding – as indicated by JUND ChIP-seq. Although, roughly 50% of the HIF binding sites that overlap an AP-1 site are not affected by AFos. This implies that the mechanisms of HIF binding to sites that overlap an AP-1 site are not identical across all sites. This is unlike the published relationship between the glucocorticoid receptor (GR) and AP-1, whereby the sole mechanism for GR binding site recruitment to sites shared with AP-1 was the maintenance of chromatin accessibility mediated by AP-1 (Biddie et al, 2011). It is possible that AP-1 is acting as a pioneer factor and subsequently maintaining accessible chromatin to enable HIF binding at a subset of shared HIF and AP-1 sites, although further investigations utilising DNaseI-,

FAIRE-, or ATAC-seq are required for determining the extent to which AP-1 and chromatin accessibility influences HIF binding.

The HIF sites that overlapped an AP-1 binding site, that were also affected by the AP-1 dominant negative, had a propensity to be associated with the positive regulation of nearby genes. Both AFos induction, and siRNAs targeting HIF-1 α and HIF-2 α , lead to a down-regulation of the nearby gene. This suggests that the binding of both HIF and AP-1 are contributing to the upregulation of these genes. Interestingly, enriched among these genes are genes which associate with cell motility and chemotaxis, thus suggesting that HIF and AP-1 may be functioning in a cooperative fashion – from the perspective of both chromatin binding and gene regulation – and subsequently contributing towards a potentially oncogenic phenotype.

An important caveat to consider is the fact that I have only performed ChIP-seq against a single AP-1 component (JUND). The different AP-1 components have been stated to have different functions (for instance, either pro- or anti-tumourigenic) depending on the context in which they are being studied (Shaulian et al, 2010). Further work exploring the binding association between HIF and other AP-1 components should be performed in the future. I am also yet to assess how the presence and absence of HIF influences the pan-genomic binding of AP-1, and this too would provide an interesting avenue for further study. Nonetheless, this data represents an important discovery as it is the first study to uncover a widespread determinant of HIF chromatin binding outside of the hypoxia response element and heterodimerisation with HIF-1 β .

6.6 Conclusion

Overall, this thesis demonstrates a significant advancement of knowledge regarding a broad range of the DNA-binding properties and DNA-binding determinants of HIF-1 and HIF-2. Of particular importance is the fixed association of HIF-1 with promoter-proximal binding sites and HIF-2 with promoter-distal, in addition to the groundbreaking discovery regarding the widespread pan-genomic cooperativity between HIF and another transcription factor in AP-1. It is likely that there are other factors, or specific histone environments, that contribute to the differential binding site selectivity of HIF-1 and HIF-2. Therefore, the data presented in this thesis provides interesting and potentially profitable new areas to explore regarding the determinants of HIF binding.

Chapter 7: Materials and methods

7.1 Cell lines

786O, MCF7 and HepG2 cell lines were purchased from American Type Culture Collection (ATCC, via LGC, U.K.). HKC8 cells were a generous gift from L. Rascusen, Baltimore, USA, and were received from K.U. Eckardt, Erlangen, Germany. RCC4 cells were a generous gift from C. H. Buys, and were later stable transfected with pcDNA3 (G418 resistant) by Ya-Min Tian (subsequently termed RCC4-VA). All parental cell lines were validated using an STR (short tandem repeats) genotyping service provided by Eurofins Genomics.

7.2 Cell culture

All cell lines were grown Dulbecco's Modified Eagle's Medium containing 100 U/ml penicillin, 100 µg/ml streptomycin and 10% fetal bovine serum (Sigma, U.K.) at 37°C and 5% CO₂ in a cell culture incubator. Cells were cultured in hypoxia using a Ruskin Invivo2 400 work station. Whilst using the Ruskin hypoxia workstation, temperature was maintained at 37°C and CO₂ tension was maintained at 5%, however O₂ tension was capable of being precisely modulated between 0.1-21% O₂.

7.3 Generic DNA techniques

7.3.1 Quantification of DNA concentration

Double-stranded DNA concentrations greater than 10 ng/µl were quantified by measuring the absorbance of the DNA solution at 260 nm using a NanoDrop® 8000 spectrophotometer (Thermo Fisher Scientific, U.K.). DNA concentrations less than 10 ng/µl were quantified using the Qubit™ 2.0 Fluorometer (Thermo

Fisher Scientific, U.K.) and the accompanying Qubit™ dsDNA HS Assay Kit.

7.3.2 Polymerase chain reaction (PCR) amplification of DNA

PCR was performed in 25 µl reactions using a Veriti™ Thermal Cycler (Applied Biosystems, U.K.). PCR reactions were performed using the Q5® High-Fidelity DNA Polymerase (NEB, U.K.) and subjected to 35 cycles of amplification. Reaction composition below.

Component	Stock	Volume
Q5 Reaction Buffer	5X	5 µl
Forward primer	10 µM	1.25 µl
Reverse primer	10 µM	1.25 µl
dNTPs	10 mM	0.5 µl
Template DNA	Variable	Variable
5X Q5 High GC Enhancer	5X	5 µl
Q5 polymerase	-	0.25 µl
H ₂ O	-	to 25 µl

7.3.3 Restriction digest of DNA

1 µg of plasmid DNA, or the entirety of a PCR reaction, were digested in 50 µl reactions for 1-4 hours at 37°C using 10-20 U of the appropriate restriction enzymes (NEB, U.K.) and the recommended reaction buffer. Alkaline phosphatase treatment to remove 5' phosphate groups was performed by adding 10 U of calf intestinal alkaline phosphatase (NEB, U.K.) for 30 minutes at 37°C. Restriction digest efficiency was verified by running the digested sample on an agarose gel. Relevant bands were excised from the agarose gel and purified as outlined in 7.3.5.

7.3.4 Gel electrophoresis

0.8-2% agarose gels were prepared by solubilising agarose in Tris/Borate/Acetate (TBE) buffer. After cooling to approximately 50°C, SYBR™ Safe DNA Gel Stain (Thermo Fisher Scientific, U.K.) was added at 1:10,000 and gels were cast. DNA samples were mixed with 6X Gel Loading Dye (NEB, U.K.) and electrophoresis was performed using a Bio-Rad gel electrophoresis chamber (Bio-Rad, U.K) at 100 volts for approximately 30 minutes. DNA sequence length was determined with reference to 1 kb or 100 bp DNA Ladder (NEB, U.K.) and were visualised using the Bio-Rad Chemidoc XRS+ system using blue light excitation (~502 nm).

7.3.5 Purification of amplified or digested DNA

PCR products or digested vectors run on agarose gels were purified using the QIAquick Gel Extraction Kit (Qiagen, U.K.) according to the manufacturer's instructions. PCR products in solution were purified using the QIAquick PCR Purification Kit (Qiagen, U.K.) according to the manufacturer's instructions. Small quantities of DNA (< 5 µg; 70 bp - 4 kb), such as those yielded during chromatin immunoprecipitation experiments, were purified using the MinElute PCR Purification Kit (Qiagen, U.K.) according to the manufacturer's instructions.

7.3.6 Ligation of DNA fragments

Ligation of digested DNA was performed using T4 Ligase (NEB, U.K.). A threefold molar excess of digested insert DNA to digested vector was used. The reaction was carried out for 1 hour at room temperature as outlined below.

Component	20 μ l Reaction
10X T4 DNA Ligase Buffer	2 μ l
Vector DNA	50 ng
Insert DNA	Variable
H ₂ O	to 20 μ l
T4 DNA Ligase	1 μ l

7.3.7 Generation of new vector constructs

Sequences to be studied were ligated into a pTRIBZ vector using the multiple cloning site. The pTRIBZ vector is a modified version of the GE Dharmacon pTRIPZ vector. Former lab member, Mathew Coleman, digested the pTRIPZ vector to remove its puromycin resistance cassette and subsequently insert a blasticidin resistance cassette. The multiple cloning site of pTRIBZ is as follows
 AgeI – AscI – SnaBI – SbfI.

The AFos sequence was amplified from a parental pCMV-500 vector after purchasing this plasmid from Addgene (Plasmid #33353). Oligonucleotide sequences used for PCR amplification of AFos may be found in section 7.9 of this Chapter. The AFos sequence was ligated into pTRIBZ after digesting the AFos amplicon and vector with AscI and SbfI high fidelity restriction enzymes (NEB, U.K.).

7.3.8 Bacterial strains, transformation and isolation of constructs

For transformation steps, NEB Turbo Competent *E.coli* (High Efficiency) cells were used according to the manufacturer's instructions.

- 1) Cells were thawed on ice
- 2) 20 μ l of cell suspension was transferred to a 1.5 ml tube and all 20 μ l of

ligation mixture was added.

- 3) Reactions were incubated on ice for 30 minutes prior to heat shocking the samples for 30 seconds in a 42°C water bath.
- 4) Samples were returned to ice for 2 minutes prior to adding 40 µl of SOC Outgrowth Medium (NEB, U.K.). The transformation reaction was then gently mixed by pipetting, and then incubated at 37°C for 1 hour.
- 5) LB agarose plates containing ampicillin (0.1 mg/ml) were pre-warmed at 37°C for 30 minutes.
- 6) After 1 hour at 37°C, transformation reactions were gently mixed by pipetting then transferred to a pre-warmed agarose plate, spread evenly across plate, and incubated overnight at 37°C.
- 7) Control ligations containing digested vector but no insert were processed in parallel to estimate background bacterial growth.
- 8) Colonies were picked from the LB agarose plates and transferred to a universal tube containing 5 ml LB medium containing ampicillin (0.1 mg/ml) and incubated in a 37°C shaking incubator at 200 rpm for approximately 16 hours.
- 9) After the 16 hour incubation 500 µl of the sample was collected and diluted 1:1 with glycerol (Sigma, U.K.). Diluted samples were subsequently stored at -80°C for long term storage of transformed bacteria. The remaining bacterial cells were collected by centrifugation of the universal tubes at 3750 rpm in a Beckman Coulter bench-top centrifuge (Allegro X-12R, Rotor SX-4750, radius: 207.8 mm).
- 10) Plasmids were isolated using plasmid purification kits such as the Qiagen QIAprep Spin Miniprep kit (Qiagen, U.K.) according to the manufacturer's

instructions.

7.3.9 DNA and RNA sequencing

‘Sanger’ sequencing of DNA was performed by Source Bioscience (Oxford, U.K.). Next generation sequencing (NGS) of DNA or RNA was performed by the Wellcome Trust Centre for Human Genetics High Throughput Genomics Core using either the HiSeq2500 or HiSeq4000 (Illumina, U.K.) instruments. All next generation sequencing was performed 75 base paired end.

7.3.10 Transfection of plasmid DNA and lentivirus production

HEK293T cells were transfected with lentiviral packaging and envelope vectors, in addition to pTRIBZ-AFos in order to produce lentivirus. 2.2×10^6 293T cells were seeded into a 10 cm dish 24 hours prior to transfection, reaching 70% confluency at the point of transfection. Two 1.5 ml tubes were used for preparing transfection reactions. 100 μ l of Opti-MEM (Gibco, U.K.) was added to the first tube, along with 3 μ g of pCMV- Δ 8.91 (lentiviral packaging), 0.7 μ g of pCMV-VSV-G (lentiviral envelope) and 3 μ g of pTRIBZ-AFos. Packaging and envelope vectors were kindly provided by Professor Kamil Kranc (Centre for Regenerative Medicine, University of Edinburgh). 400 μ l of Opti-MEM was added to tube 2 along with 21 μ g of polyethylenimine (PEI) transfection reagent (Sigma, U.K.). Both tubes were vortexed and tube one was added to tube 2. Tube 2 was then mixed by gently flicking. The transfection mixture was then left to incubate at room temperature for 20 minutes prior to adding the mixture to the 293T cells in a drop-wise manner. Media was changed on the 293T cells 24 hours after transfection. Lentivirus-containing media was subsequently harvested 48 hours

post-transfection. Harvested media was syringe filtered (0.45 μm) prior to snap-freezing on dry ice and storage at -80°C .

7.3.11 Lentiviral transduction using plasmid DNA

Lentivirus-containing media is thawed on ice and mixed 1:1 with DMEM containing 10% FCS. Transduction enhancer, polybrene (Sigma, U.K.) was added to a final concentration of 4 $\mu\text{g/ml}$. The 50% diluted lentivirus-containing media is then added directly to 70% confluent RCC4-VA cells. After 24 hours, media is removed and fresh DMEM containing 10% FCS is applied. After a further 24 hours, blasticidin is added to a final concentration of 30 $\mu\text{g/ml}$ to select for positively transduced RCC4 pTRIBZ-AFos cells.

7.4 Protein techniques

7.4.1 Isolation of whole cell extracts

For isolation of whole cell extracts, cells were cultured to sub-confluency in 6- or 12-well dishes. After desired cell culture duration, medium was removed and cells were washed twice with ice-cold PBS. Cells were scraped into 120 μl (for 6-well dishes) or 80 μl (12-well) Urea-Tris Buffer (see 7.10.1) supplemented with cOmplete™ protease inhibitor (Roche, U.K.) at 1X. Lysates were incubated on ice in 1.5 ml tubes for 10 minutes prior to sonication for 20 seconds on medium power using a 4°C waterbath Bioruptor® sonicator (Diagenode, U.K.). Sonicated lysates were then centrifuged for 10 minutes at 13,500 rpm using an Eppendorf™ benchtop microcentrifuge (Model 5415R), chilled to 4°C, to remove any insoluble material. Supernatant is collected and subsequently mixed 1:1 with 2X sample buffer (7.10.17) and then stored at -20°C .

7.4.2 Immunoblotting

SDS-PAGE (sodium dodecyl sulphate polyacrylamide gel electrophoresis) was used to resolve proteins of interest. Resolving acrylamide gels were prepared using SDS-PAGE resolving gel buffer (composition in Methods section 7.10.5) and 30% acrylamide/Bis-acrylamide solution (Sigma, U.K.) to final concentrations of between 8-15%, depending on the molecular weight of the proteins of interest. Stacking gels were prepared using SDS-PAGE stacking gel buffer (7.10.6) and 30% acrylamide/Bis-acrylamide solution to a final concentration of 3.9%. Polymerisation of gels was achieved by the addition of ammonium persulphate (0.1% w/v final, Sigma, U.K.) and N,N,N',N'-tetramethylethylenediamine (0.1% w/v final, Sigma, U.K.).

- 1) SDS-PAGE was performed using a Mini-Protean apparatus (Bio-Rad, U.K.). Up to 12.5 μ l of sample was added to relevant wells of the SDS-PAGE gels, and 4 μ l of protein ladder (PageRuler Plus, Fermentas, U.K.) was added to a single well on the same gel.
- 2) Electrophoresis was carried out in SDS-PAGE running buffer (1X) at a constant voltage of 120 volts for 1-2 hours.
- 3) Proteins were transferred onto polyvinylidene fluoride (PVDF) membrane (Millipore, U.K.) using the Mini-Transblot apparatus (Bio-Rad, U.K.). PVDF membrane was activated for at least 1 minute in 100% methanol (Sigma, U.K.), then equilibrated in for at least 1 minute in western blot transfer buffer (1X) (7.10.4). Western blot transfer was carried out at 320 mA for 25 minutes with ice cooling.
- 4) Membranes were then blocked in western blot blocking buffer (7.10.8 containing 5% skim milk powder) for 1 hour at room temperature.

- 5) Membranes were subsequently incubated with a specific primary antibody (diluted in western blot blocking buffer) overnight at 4°C.
- 6) PBS-T (7.10.8) was used to wash the membranes following incubation with the primary antibody. Membranes are washed with multiple applications of PBS-T at room temperature.
- 7) HRP-conjugated secondary antibody (anti-mouse IgG or anti-rabbit IgG, Dako, U.K.) diluted in western blot blocking buffer at 1:2000 was then added to the membranes and incubated for 1 hour at room temperature.
- 8) Membranes were then washed with multiple applications of PBS-T at room temperature.
- 9) Immunodetection was performed by visualising the HRP activity by using the chemiluminescence reagent, SuperSignal West Dura Extended Duration Substrate (Thermo Fisher Scientific, U.K.). Immunoblots were visualised using the Bio-Rad Chemidoc XRS+ system.

7.5 RNA techniques

7.5.1 RNA isolation

Total RNA was extracted from sub-confluent adherent cells cultured in 10 cm dishes using the RNeasy Micro Kit (Qiagen, U.K.) according to the manufacturer's instructions. Eluted RNA was quantified using a NanoDrop® 8000 spectrophotometer (Thermo Fisher Scientific, U.K.) and snap frozen prior to storage at -80°C.

7.5.2 RNA quality control

RNA integrity was assessed using an Agilent 2200 TapeStation and the High

Sensitivity or Standard RNA ScreenTape and Buffer (Agilent, U.K.). RNA Integrity Numbers (RIN) provided an indication of the quality of RNA. Only samples with RIN values of >8 were deemed to be of sufficient quality for submission for high throughput sequencing.

7.6 Chromatin immunoprecipitation (ChIP)

7.6.1 Procedure for isolating ChIP DNA

Cells cultured for ChIP were grown in 15 cm dishes and cultured to sub-confluency. Between 1.5 and 2 x 15 cm dishes were used per each individual ChIP assay, which equates to roughly 25 million cells per ChIP. VHL-defective cells were harvested in normoxia, whereas VHL-competent cells were cultured in 0.5% O₂ for 16 hours prior to harvesting.

- 1) Cells were cross-linked by adding formaldehyde to the cell culture medium in each dish to a final concentration of 1%. Dishes were sat on ice for 12 minutes and briefly rocked every minute for of the duration of the 12 minutes. For cells being cultured in hypoxia workstations, formaldehyde was added to the medium whilst the cells were still in the workstation to avoid re-oxygenation of cells and subsequent degradation of HIF- α . Cells were quickly transferred out of the station and on to ice after formaldehyde addition.
- 2) Formaldehyde was quenched by adding 1 M glycine in PBS to the cell culture medium to a final concentration of 0.125 M for a further 10 minutes. Dishes were rocked at room temperature for the duration of the 10-minute incubation.
- 3) Cells were washed three times with ice-cold PBS, and subsequently

scraped into 5 ml of ice-cold PBS and transferred into a 50 ml tube.

- 4) Cells were pelleted by centrifugation at 800 rpm at 4°C for 10 minutes (Allegro X-12R, Rotor SX-4750, radius: 207.8 mm).
- 5) Supernatant was discarded and cell pellets were resuspended in 500 μ l in ChIP SDS lysis buffer per 15 cm plate of pelleted cells. Lysates were incubated on ice for 10 minutes.
- 6) Lysates were then diluted 1:1 using ChIP-dilution buffer and each 1 ml of lysate was transferred to individual 15 ml TPX sonication tubes (Diagenode, U.K.).
- 7) Lysates were sonicated for 15 minutes using 4°C waterbath Bioruptor® sonicator (Diagenode, U.K.). Sonication was conducted using the high power setting whilst cycling 15 seconds on/off.
- 8) Sonicated cell lysates were then transferred to 1.5 ml tubes and centrifuged at 13500 rpm at 4°C to remove any insoluble material. An aliquot, corresponding to 2% of an individual ChIP sample, was retrieved to serve as ChIP input material. Input lysate was stored at 4°C overnight. Sonicated chromatin was then transferred to 15 ml tubes and diluted 1:8 in ChIP-dilution buffer. Diluted chromatin was then aliquoted for each ChIP into 15 ml tubes accordingly.
- 9) Protein A agarose beads (Millipore, U.K.) were then washed three times with ChIP-dilution buffer. 100 μ l of washed beads was added to each ChIP lysate aliquot for “pre-clearing”. Diluted chromatin and washed beads were incubated together for 1 hour at 4°C on an end-over-end rotator.
- 10) Samples were then centrifuged at 380 x g (Allegro X-12R, Rotor SX-

4750, radius: 207.8 mm) to retrieve the “pre-cleared” supernatant. Antibodies were then added to the chromatin for immunoprecipitation overnight on an end-over-end rotator at 4°C. 10 µl of in-house antibodies were used per ChIP (PM-14 for HIF-1α; PM-9 for HIF-2α), whereas between 10 - 20 µl was used depending on the commercial antibody. During assays intended for ChIP-qPCR, rabbit serum was used for a negative control ChIP.

- 11) Protein A agarose beads were washed three times in ChIP-dilution buffer, then 80 µl was added to each ChIP for capturing immunoaffinity complexes. Beads were incubated with the samples for 2 hours prior to centrifugation for 8 minutes at 380 x g (Allegro X-12R, Rotor SX-4750, radius: 207.8 mm) to collect the bead-captured immunoaffinity complexes.
- 12) Supernatant was discarded and beads were resuspended in 1 ml of ice cold ChIP low salt buffer and transferred to 1.5 ml Eppendorf® DNA LoBind tubes
- 13) Samples were incubated at 4°C for 5 minutes on an end-over-end rotator and then centrifuged at 380 x g for 3 minutes at 4°C.
- 14) Supernatant was removed and bead washing steps were repeated using 1 ml ice cold ChIP high salt buffer, ChIP LiCl buffer, and TE buffer (twice).
- 15) Following the last TE wash, all residual TE buffer was removed using a gel loading tip. 120 µl of ChIP elution buffer was then added and samples were agitated on an Eppendorf Thermomixer® F1.5 at 1400 rpm at room temperature for 15 minutes.

- 16) Samples were centrifuged at 380 x *g* for 3 minutes at room temperature, and the supernatants were transferred to fresh 1.5 ml DNA LoBind tubes.
- 17) The elution step was repeated by applying a fresh 120 µl of ChIP elution buffer to the beads.
- 18) ChIP elution buffer was added to input samples up to 240 µl. 12.5 µl of 5 M NaCl was added to 240 µl of eluate in addition to each input sample.
- 19) Samples were then incubated at 65°C on an Eppendorf Thermomixer® F1.5 at 1400 rpm overnight.
- 20) The next day, temperature on the thermomixer was reduced to 45°C, and 2 µl of 20 mg/ml proteinase K (Roche, U.K.) was added to each sample, whilst the samples remain mixing at 1400 rpm for 4 hours.
- 21) Temperature on the thermomixer was then reduced to 37°C and 1 µl of 10 mg/ml RNaseA (DNase and protease-free; Thermo Fisher Scientific, U.K.) was added to each sample, whilst the samples remain mixing at 1400 rpm for 30 minutes.
- 22) DNA was then purified using the MinElute PCR Purification Kit (Qiagen, U.K.).

7.6.2 DNA quality control

Sonication-induced DNA shearing, and subsequent size distribution of DNA fragments was assessed using an Agilent 2200 TapeStation and the High Sensitivity DNA ScreenTape and Buffer (Agilent, U.K.). Median DNA fragment size between 300 - 500 bp was deemed optimal and of sufficient quality for

submission for high throughput sequencing.

7.6.3 Quantitative analysis of ChIP DNA using SYBR green assay

ChIP enrichment of specific DNA interacting proteins was determined by qPCR using the StepOnePlus (Applied Biosystems, U.K.), and subsequently quantified using the percentage input method.

7.7 Bioinformatics tools

7.7.1 ChIP-seq pipeline

Initial analysis of raw ChIP-seq data was achieved through the use of a ChIP-seq analysis pipeline developed by former member of the Prof. Sir Peter Ratcliffe's group, Rafik Salama (full details of this can be found here: Salama et al., 2015). In brief, Illumina adaptor sequences were trimmed using Trimgalore (0.3.3) and reads were aligned to Genome Reference Consortium GRCh37 (hg19) using Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2009). Low quality mapped reads (MapQ < 15) were removed using SAMtools (0.1.19) (Li et al., 2009a), and all reads mapping to ENCODE Duke University black listed regions (<http://hgwdev.cse.ucsc.edu/cgi-bin/hgFileUi?db=hg19&g=wgEncodeMapability>) were discarded using BEDTools (Dale et al., 2011). Duplicate reads were marked for exclusion using Picard tools (1.106) (<http://broadinstitute.github.io/picard/>). ChIP-seq peaks and peak summits (1 bp coordinate at which a ChIP-seq peak is at its highest) were identified using MACS (Model-based analysis of ChIP-seq) (Zhang et al., 2008b).

7.7.2 BEDTools

BEDTools intersect (Dale et al., 2011) was used to identify ChIP-seq peaks that are overlapping between one .bed file (text file comprising ChIP-seq peak coordinates, as defined by their chromosome number, start base and end base) and another. The '-wa' option was used to write only the original entry for .bed file A within the output .bed file. The minimum overlap required for a region to qualify as intersecting was 1 bp.

BEDTools subtract was used to identify ChIP-seq peaks that are not overlapping between one .bed file and another. The '-A' option was used to ensure that if any peak in .bed file A overlaps a peak in .bed file B, the entire overlapping peak in .bed file A is removed from the output file.

BEDTools multicov was used to map .bam (read alignments) file reads to regions provided in the form of a .bed file.

BEDTools getfasta was used to extract nucleotide sequences from an hg19 fasta file (text file comprising nucleotide sequences) for each of the regions defined within a provided .bed file.

The complete list BEDTools commands and options can be viewed here - <https://bedtools.readthedocs.io/en/latest/content/bedtools-suite.html>

7.7.3 SAMtools

SAMtools merge was used to merge sorted alignment .bam files into a single .bam output file (Li et al., 2009a). SAMtools index was used to generate index files (.bai) from sorted alignment files to permit further analysis. SAMtools flagstat was used to obtain simple alignment statistics. The totality of SAMtools commands and options can be viewed here - <http://www.htslib.org/doc/samtools->

7.7.4 Motif analysis and MEME-suite

Fasta files comprising sequences flanking each peak summit (± 150 bp) were repeat-masked using RepeatMasker 4.0.3 (www.repeatmasker.org). AME (Analysis of Motif Enrichment) - from within the MEME-suite of motif-based analysis tools - was used to identify known motifs enriched within provided repeat-masked sequences, as compared with shuffled versions of the provided sequences (McLeay and Bailey, 2010). The motif library used for analysis was the JASPAR Core (2016) Vertebrates library (<http://jaspar.genereg.net/>).

7.7.5 Heatmaps and ngs.plot

Binding site heatmaps and average profile line plots were generated using ngs.plot (Shen et al., 2014). ngs.plot parameters were: -GO none (for ranking (y-axis) of genes or regions in heatmaps), -SE 0 (standard error), -FL 50 (fragment length), -MQ 15 (read mapping quality cut-off according to Phred-scale), -L 5000 (flanking regions size). A full list of parameters may be found here <https://github.com/shenlab-sinai/ngsplot/wiki/ProgramArguments101>

7.7.6 Gene set enrichment analysis (GSEA)

GSEA was performed using 10000 permutations, a weighted enrichment score and pre-ranked genes (Subramanian et al., 2005). Differential expression and significance according to DESeq2 and fold difference between expression of genes in normoxia and 0.5% O₂ were used to rank genes (Xiao et al., 2014). The equation is as follows:

$$x = -\log_{10}[\text{adjusted } p \text{ value}] * \log_2[\text{fold change}]$$

7.7.7 Differential binding analysis

Differential binding analysis of sites determined via ChIP-seq experiments was performed using the R package, DiffBind (Ross-Innes et al., 2012).

7.7.8 RNA-seq pipeline

Initial analysis of raw RNA-seq data was achieved through the use of a RNA-seq analysis pipeline developed by former member of the Prof. Sir Peter Ratcliffe's group, Rafik Salama (Salama et al., 2015). In brief, Illumina adaptor sequences were trimmed using Trimgalore (0.3.3). RNA-seq reads were aligned to GRCh37 using Tophat 2.0.8b (<http://ccb.jhu.edu/software/tophat/index.shtml>) and bowtie (<http://bowtie-bio.sourceforge.net/index.shtml>), and non-uniquely mapping reads were excluded from analysis using SAMtools (0.1.19) (Li et al., 2009a). The totality of read counts for each UCSC defined gene were extracted using HTSeq (0.5.4p3) (Anders et al., 2015) with 'intersect strict' mode and significantly regulated genes were determined through the use of DESeq2 (Love et al., 2014).

7.8 Antibodies

Name	Source	Cat #	Host	Application
HIF-1 β	Novus Biologicals	NB 100-110	Rabbit antiserum	ChIP, western
HIF-1 α	In-house	PM-14	Rabbit antiserum	ChIP
HIF-2 α	In-house	PM-9	Rabbit antiserum	ChIP
JUND	Santa Cruz Biotechnology	sc-74	Purified rabbit polyclonal	ChIP
HIF-1 α	BD Bioscience	610959	Mouse monoclonal	Western
HIF-2 α	In-house	190b	Mouse monoclonal	Western
FLAG (HRP-conjugated)	Sigma	A8592	Mouse monoclonal	Western
Rabbit serum	Dako	X090210-8	Rabbit serum	ChIP

7.9 Oligonucleotides

Name	Forward or reverse	Sequence	Application
<i>ACTG1 (F)</i>	Forward	AGATGTGGATTAGCAAGCAGG	ChIP-qPCR
<i>ACTG1 (R)</i>	Reverse	GCTTATTCCAGTTTCGTGAGGC	ChIP-qPCR
<i>SEMA4B (F)</i>	Forward	CCCAGAGCTGCAATGTGTAC	ChIP-qPCR
<i>SEMA4B (R)</i>	Reverse	GTAGGGCTGTGCAAACCTCC	ChIP-qPCR
<i>NDRG1 (F)</i>	Forward	TCCCTCCCAATCTCTCTTCTT	ChIP-qPCR
<i>NDRG1 (R)</i>	Reverse	CACCATCAGDACAGCAAACCTAC	ChIP-qPCR
AFos Ascl (F)	Forward	TTCATGGCGCGCCATGGACTAC AAGGACGACGATGACAAG	Cloning AFos
AFos SbfI (R)	Reverse	TTCATCCTGCAGGGAATTAATC AGGGATCTTGCAGGCAGG	Cloning AFos

7.10 Solutions and buffers

7.10.1 Urea-Tris cell lysis buffer

Reagent	Final
Urea	9 M
Tris-HCl pH 7.5	75 mM
β -mercaptoethanol	0.15 M

7.10.2 Whole cell extract sample buffer (2x)

Reagent	Final
Urea	6 M
Tris-HCl pH 7.5	17 mM
SDS	3.3%
Bromophenol blue	0.01%
β -mercaptoethanol	0.07 M

7.10.3 Western blot running buffer (10x)

Reagent	Final
Tris	0.25 M
Glycine	1.9 M
SDS	1%

7.10.4 Western blot transfer buffer (20x)

Reagent	Final
Tris	0.2 M
Glycine	2 M

7.10.5 SDS-PAGE resolving gel buffer

Reagent	Final
Tris-HCl pH 8.8	1.5 M
SDS	0.4%

7.10.6 SDS-PAGE stacking gel buffer

Reagent	Final
Tris-HCl pH 6.8	1.5 M
SDS	0.4%

7.10.7 Phosphate buffered saline (PBS)

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

7.10.8 PBS-T

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

7.10.9 ChIP SDS lysis buffer

Reagent	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 protease inhibitor	50x	0.02 ml	1x
H ₂ O		0.86 ml	

7.10.10 ChIP dilution buffer

Reagent	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 protease inhibitor	50x	1 ml	1x
H ₂ O		45.8 ml	

7.10.11 ChIP low salt buffer

Reagent	Stock	Volume	Final
SDS	20%	0.005 ml	0.1%
EDTA	0.5 M	0.004 ml	2 mM
Tris pH 8.1	1 M	0.002 ml	20 mM
Triton X-100	100%	0.01 ml	1%
NaCl	5 M	0.03 ml	150 mM
H ₂ O		0.931 ml	

7.10.12 **ChIP high salt buffer**

Reagent	Stock	Volume	Final
SDS	20%	0.005 ml	0.1%
EDTA	0.5 M	0.004 ml	2 mM
Tris pH 8.1	1 M	0.002 ml	20 mM
Triton X-100	100%	0.01 ml	1%
NaCl	5 M	0.1 ml	500 mM
H ₂ O		0.861 ml	

7.10.13 **ChIP lithium chloride buffer**

Reagent	Stock	Volume	Final
LiCl	8 M	0.032 ml	0.25 M
Igepal CA-630 NP40	100%	0.01 ml	1%
Sodium deoxycholate		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 ₂ O		0.946 ml	

7.10.14 **TE buffer**

Reagent	Stock	Volume	Final
Tris pH 8.0	1 M	0.01 ml	10 mM
EDTA	0.5 M	0.002 ml	1 mM
H ₂ O		0.988 ml	

7.10.15 ChIP elution buffer

Reagent	Stock	Volume	Final
NaHCO ₃		0.0841 g	0.1 M
SDS	20%	0.5 ml	1%
H ₂ O		9.5 ml	

Chapter 8: Appendix

8.1 Accession numbers for published NGS raw data

Cell line	Experiment	GEO accession
7860	FAIRE-seq	GSM1011120
7860	H3K27ac ChIP-seq	GSM2067534
HepG2	DNaseI-seq	GSM2400287
HepG2	H3K27ac ChIP-seq	GSM733743
RPTEC	DNaseI-seq	GSM736543
RPTEC	H3K4me3 ChIP-seq	GSM945216
HepG2	GABPA ChIP-seq	GSM803343
HepG2	YY1 ChIP-seq	GSM803381
HepG2	JUND ChIP-seq	GSM803500
HepG2	BHLHE40 ChIP-seq	GSM803483
HepG2	H3K27me3 ChIP-seq	GSM733754
HepG2	H3K9ac ChIP-seq	GSM733648
HepG2	H3K4me1 ChIP-seq	GSM798321
HepG2	H3K4me2 ChIP-seq	GSM733693
HepG2	H3K4me3 ChIP-seq	GSM733737
HepG2	H3K79me2 ChIP-seq	GSM733641
HepG2	H3K36me3 ChIP-seq	GSM733685

8.2 Accession numbers for ENCODE .bed files

Cell line	Experiment	ENCODE .bed accession
HepG2	CTCF ChIP-seq (1)	ENCFF002CDZ
HepG2	EZH2 ChIP-seq	ENCFF002CEA
HepG2	ATF3 ChIP-seq	ENCFF002CKB
HepG2	BHLHE40 ChIP-seq (1)	ENCFF002CKC
HepG2	CEBPB ChIP-seq (1)	ENCFF002CKD
HepG2	CEBPD ChIP-seq	ENCFF002CKE
HepG2	CTCF ChIP-seq (2)	ENCFF002CKF
HepG2	ELF1 ChIP-seq	ENCFF002CKG
HepG2	FOSL2 ChIP-seq	ENCFF002CKH
HepG2	FOXA1 ChIP-seq (1)	ENCFF002CKI
HepG2	FOXA1 ChIP-seq (2)	ENCFF002CKJ
HepG2	FOXA2 ChIP-seq	ENCFF002CKK
HepG2	GABPA ChIP-seq	ENCFF002CKL
HepG2	HDAC2 ChIP-seq	ENCFF002CKM
HepG2	HNF4A ChIP-seq	ENCFF002CKN
HepG2	HNF4G ChIP-seq	ENCFF002CKO
HepG2	JUND ChIP-seq (1)	ENCFF002CKP
HepG2	MBD4 ChIP-seq	ENCFF002CKQ
HepG2	MYBL2 ChIP-seq	ENCFF002CKR
HepG2	NFIC ChIP-seq	ENCFF002CKS

Cell line	Experiment	ENCODE .bed accession
HepG2	REST ChIP-seq (1)	ENCFF002CKT
HepG2	REST ChIP-seq (2)	ENCFF002CKU
HepG2	EP300 ChIP-seq (1)	ENCFF002CKV
HepG2	POLR2A P-S5 ChIP-seq	ENCFF002CKW
HepG2	POLR2A ChIP-seq (1)	ENCFF002CKX
HepG2	RAD21 ChIP-seq (1)	ENCFF002CKY
HepG2	RXRA ChIP-seq	ENCFF002CKZ
HepG2	SIN3A ChIP-seq	ENCFF002CLA
HepG2	SP1 ChIP-seq	ENCFF002CLB
HepG2	SP2 ChIP-seq	ENCFF002CLC
HepG2	SRF ChIP-seq	ENCFF002CLD
HepG2	TAF1 ChIP-seq	ENCFF002CLE
HepG2	TCF12 ChIP-seq	ENCFF002CLF
HepG2	TEAD4 ChIP-seq	ENCFF002CLG
HepG2	USF1 ChIP-seq	ENCFF002CLH
HepG2	YY1 ChIP-seq	ENCFF002CLI
HepG2	ZBTB33 ChIP-seq	ENCFF002CLJ
HepG2	ZBTB7A ChIP-seq	ENCFF002CLK
HepG2	ARID3A ChIP-seq	ENCFF002CTS
HepG2	BHLHE40 ChIP-seq (2)	ENCFF002CTT

Cell line	Experiment	ENCODE .bed accession
HepG2	BRCA1 ChIP-seq	ENCFF002CTU
HepG2	CEBPB ChIP-seq (2)	ENCFF002CTV
HepG2	CEBPB ChIP-seq (3)	ENCFF002CTW
HepG2	CHD2 ChIP-seq	ENCFF002CTX
HepG2	JUN ChIP-seq	ENCFF002CTY
HepG2	ESRRA ChIP-seq	ENCFF002CUA
HepG2	NR3C1 ChIP-seq	ENCFF002CUB
HepG2	HNF4A ChIP-seq	ENCFF002CUC
HepG2	HSF1 ChIP-seq	ENCFF002CUD
HepG2	IRF3 ChIP-seq	ENCFF002CUE
HepG2	JUND ChIP-seq (2)	ENCFF002CUF
HepG2	MAFF ChIP-seq	ENCFF002CUG
HepG2	MAFK ChIP-seq (1)	ENCFF002CUH
HepG2	MAFK ChIP-seq (2)	ENCFF002CUI
HepG2	MAX ChIP-seq	ENCFF002CUJ
HepG2	MXI1 ChIP-seq	ENCFF002CUL
HepG2	NRF1 ChIP-seq	ENCFF002CUM
HepG2	EP300 ChIP-seq (2)	ENCFF002CUN
HepG2	POLR2A ChIP-seq (2)	ENCFF002CUP
HepG2	POLR2A ChIP-seq (3)	ENCFF002CUQ

Cell line	Experiment	ENCODE .bed accession
HepG2	POLR2A P-S2 ChIP-seq	ENCFF002CUR
HepG2	RAD21 ChIP-seq (2)	ENCFF002CUS
HepG2	RFX5 ChIP-seq	ENCFF002CUT
HepG2	SMC3 ChIP-seq	ENCFF002CUU
HepG2	SREBF1 ChIP-seq	ENCFF002CUV
HepG2	TBP ChIP-seq	ENCFF002CUW
HepG2	TCF7L2 ChIP-seq	ENCFF002CUX
HepG2	NR2C2 ChIP-seq	ENCFF002CUY
HepG2	USF2 ChIP-seq	ENCFF002CUZ
HepG2	ZNF274 ChIP-seq	ENCFF002CVA
HepG2	MYC ChIP-seq	ENCFF002DAW
HepG2	CTCF ChIP-seq (3)	ENCFF002DAX
HepG2	POLR2A ChIP-seq (4)	ENCFF002DAY
HepG2	CTC4 ChIP-seq (4)	ENCFF002DCX
HepG2	CBX1 ChIP-seq	ENCFF321OLU
HepG2	ZHX2 ChIP-seq	ENCFF414AVG
HepG2	HCF1 ChIP-seq	ENCFF810BIV

8.3 Non-canonical HIF- α peak coordinates

Cell line	Factor	Chr	Start	End
HepG2	HIF-2 α	chr22	41222928	41223494
HepG2	HIF-2 α	chr1	234579947	234580540
RCC4	HIF-2 α	chr4	102874870	102875759
RCC4	HIF-2 α	chr4	73501946	73502869
RCC4	HIF-2 α	chr4	155828718	155830040
RCC4	HIF-2 α	chr4	165125374	165126181
RCC4	HIF-2 α	chr4	26332257	26333118
RCC4	HIF-2 α	chr14	25579947	25580866
HKC8	HIF-2 α	chr1	246598560	246599562
HKC8	HIF-2 α	chr11	17192745	17193494
HKC8	HIF-2 α	chr10	25051610	25052560
RCC4	HIF-1 α	chr4	26332206	26333191

8.4 ChIP-seq mapped read counts

Cell line	ChIP-seq	Condition	Mapped reads
786O pRRL	HIF-2 α	Normoxic	2,837,959
786O pRRL	HIF-1 β	Normoxic	22,482,463
786O pRRL HIF-1 α	HIF-1 α	Normoxic	7,435,627
786O pRRL HIF-1 α	HIF-2 α	Normoxic	2,175,628
786O pRRL HIF-1 α	HIF-1 β	Noxmoxic	2,175,628
786O pRRL HIF-2 α	HIF-2 α	Normoxic	9,106,516
786O pRRL HIF-2 α	HIF-1 β	Normoxic	2,837,959
HepG2	HIF-1 α (1)	16 hours 0.5% O ₂	36,853,640
HepG2	HIF-1 α (2)	16 hours 0.5% O ₂	32,132,003
HepG2	HIF-2 α (1)	16 hours 0.5% O ₂	32,055,373
HepG2	HIF-2 α (2)	16 hours 0.5% O ₂	33,649,440
HepG2	HIF-1 β (1)	16 hours 0.5% O ₂	39,097,491
HepG2	HIF-1 β (2)	16 hours 0.5% O ₂	29,670,750
RCC4	HIF-1 α (1)	Normoxic	38,378,304
RCC4	HIF-1 α (2)	Normoxic	42,673,393

Cell line	ChIP-seq	Condition	Mapped reads
RCC4	HIF-2 α (1)	Normoxic	37,339,895
RCC4	HIF-2 α (2)	Normoxic	42,195,677
HKC8 WT	HIF-1 α (1)	16 hours 0.5% O ₂	49,249,255
HKC8 WT	HIF-1 α (2)	16 hours 0.5% O ₂	49,491,323
HKC8 WT	HIF-2 α (1)	16 hours 0.5% O ₂	45,991,172
HKC8 WT	HIF-2 α (2)	16 hours 0.5% O ₂	61,312,804
HKC8 WT	HIF-1 β (1)	16 hours 0.5% O ₂	47,824,526
HKC8 WT	HIF-1 β (2)	16 hours 0.5% O ₂	60,804,433
HKC8 -1 α KO	HIF-2 α (1)	16 hours 0.5% O ₂	50,857,968
HKC8 -1 α KO	HIF-2 α (2)	16 hours 0.5% O ₂	53,247,936
HKC8 -1 α KO	HIF-1 β (1)	16 hours 0.5% O ₂	54,963,388
HKC8 -1 α KO	HIF-1 β (2)	16 hours 0.5% O ₂	49,153,339
HKC8 -2 α KO	HIF-1 α (1)	16 hours 0.5% O ₂	52,292,525
HKC8 -2 α KO	HIF-1 α (2)	16 hours 0.5% O ₂	49,141,302
HKC8 -2 α KO	HIF-1 β (1)	16 hours 0.5% O ₂	49,477,258
HKC8 -2 α KO	HIF-1 β (2)	16 hours 0.5% O ₂	49,402,204

Cell line	ChIP-seq	Condition	Mapped reads
RCC4 pTRIBZ-AFos	JUND (1)	Vehicle	51,614,620
RCC4 pTRIBZ-AFos	JUND (2)	Vehicle	54,569,399
RCC4 pTRIBZ-AFos	HIF-1 α (1)	Vehicle	51,956,430
RCC4 pTRIBZ-AFos	HIF-1 α (2)	Vehicle	47,638,750
RCC4 pTRIBZ-AFos	HIF-2 α (1)	Vehicle	56,472,896
RCC4 pTRIBZ-AFos	HIF-2 α (2)	Vehicle	56,305,389
RCC4 pTRIBZ-AFos	HIF-1 β (1)	Vehicle	60,304,396
RCC4 pTRIBZ-AFos	HIF-1 β (2)	Vehicle	35,861,488
RCC4 pTRIBZ-AFos	JUND (1)	48 hours dox (1 μ g/ml)	47,216,955
RCC4 pTRIBZ-AFos	JUND (2)	48 hours dox (1 μ g/ml)	53,436,154
RCC4 pTRIBZ-AFos	HIF-1 α (1)	48 hours dox (1 μ g/ml)	58,116,146
RCC4 pTRIBZ-AFos	HIF-1 α (2)	48 hours dox (1 μ g/ml)	57,333,820
RCC4 pTRIBZ-AFos	HIF-2 α (1)	48 hours dox (1 μ g/ml)	55,016,335
RCC4 pTRIBZ-AFos	HIF-2 α (2)	48 hours dox (1 μ g/ml)	44,315,354
RCC4 pTRIBZ-AFos	HIF-1 β (1)	48 hours dox (1 μ g/ml)	51,160,565
RCC4 pTRIBZ-AFos	HIF-1 β (2)	48 hours dox (1 μ g/ml)	59,981,498

8.5 AME analysis and significantly enriched motifs within specified sets of genomic regions

8.5.1 Sites preferentially bound by HIF-1 in HepG2

Database ID	Transcription Factor	Adjusted <i>p</i> -value
MA0259.1	ARNT::HIF1A	2.32E-197
MA0079.3	SP1	4.62E-79
MA0516.1	SP2	8.78E-76
MA0649.1	HEY2	7.02E-73
MA0599.1	KLF5	7.97E-73
MA0528.1	ZNF263	4.11E-67
MA0004.1	Arnt	3.96E-63
MA0823.1	HEY1	3.53E-52
MA0616.1	Hes2	1.00E-51
MA0162.2	EGR1	4.32E-51
MA0039.2	Klf4	5.75E-50
MA0746.1	SP3	4.35E-46
MA0626.1	Npas2	5.38E-46
MA0136.2	ELF5	2.56E-45
MA0862.1	GMEB2	1.08E-43
MA0663.1	MLX	6.81E-43
MA0617.1	Id2	7.69E-43
MA0622.1	Mlxip	6.02E-42
MA0493.1	Klf1	1.89E-39
MA0685.1	SP4	6.27E-38
MA0741.1	KLF16	8.56E-37
MA0645.1	ETV6	1.45E-35
MA0747.1	SP8	1.02E-34
MA0464.2	BHLHE40	3.27E-33
MA0076.2	ELK4	4.70E-31
MA1099.1	Hes1	8.04E-31
MA0104.3	Mycn	4.38E-30
MA0152.1	NFATC2	5.76E-30
MA0636.1	BHLHE41	1.44E-29
MA0825.1	MNT	1.02E-28
MA0062.2	Gabpa	2.88E-28
MA0740.1	KLF14	1.55E-27
MA0664.1	MLXIPL	3.34E-27
MA0472.2	EGR2	3.39E-26
MA0058.3	MAX	3.60E-26
MA0603.1	Arntl	3.63E-26
MA0598.2	EHF	1.33E-25
MA0149.1	EWSR1-FLI1	1.98E-23
MA0471.1	E2F6	6.21E-23
MA0733.1	EGR4	1.44E-22
MA0606.1	NFAT5	2.90E-22
MA0641.1	ELF4	3.37E-22
MA0473.2	ELF1	4.04E-22
MA0624.1	NFATC1	4.71E-22
MA0762.1	ETV2	7.47E-22
MA0819.1	CLOCK	8.18E-22
MA0625.1	NFATC3	1.81E-21
MA0098.3	ETS1	2.01E-21
MA0638.1	CREB3	3.74E-21
MA0508.1	PRDM1	1.04E-20
MA0640.1	ELF3	3.15E-20
MA0059.1	MAX::MYC	9.70E-20
MA0742.1	Klf12	1.84E-19
MA0506.1	NRF1	2.32E-19
MA0604.1	Atf1	3.03E-19
MA0147.2	Myc	3.73E-18
MA0050.2	IRF1	5.03E-18
MA0474.2	ERG	8.91E-18
MA0475.2	FLI1	9.82E-18
MA0156.2	FEV	3.95E-16
MA0732.1	EGR3	7.17E-16
MA0764.1	ETV4	1.31E-14
MA0081.1	SP1B	1.89E-14
MA0517.1	STAT1::STAT2	3.99E-14
MA0760.1	ERF	4.20E-14
MA0481.1	FOXP1	5.54E-14
MA0526.1	USF2	6.10E-14
MA0844.1	XBP1	7.38E-14
MA0608.1	Creb3l2	1.88E-13
MA0822.1	HES7	2.04E-13
MA0845.1	FOXB1	5.18E-13
MA0761.1	ETV1	6.25E-13
MA0759.1	ELK3	7.73E-13
MA0032.2	FOXC1	1.12E-12
MA0144.2	STAT3	1.43E-12
MA0686.1	SPDEF	5.36E-12
MA0839.1	CREB3L1	1.93E-11
MA0609.1	Crem	3.15E-11
MA0593.1	FOXP2	3.42E-11
MA0470.1	E2F4	5.07E-11
MA0765.1	ETV5	1.27E-10
MA0131.2	HINFP	1.37E-10
MA0080.4	SP1	1.76E-10
MA0846.1	FOXC2	1.92E-10
MA0615.1	Gmeb1	2.73E-10
MA0852.1	Foxk1	1.66E-09
MA0006.1	Ahr::Arnt	2.74E-09
MA0073.1	RREB1	3.72E-09
MA0821.1	HES5	4.96E-09
MA0090.2	TEAD1	7.82E-09
MA0632.1	Tcf15	8.97E-09
MA0763.1	ETV3	9.08E-09
MA0851.1	Foxj3	1.04E-08
MA0030.1	FOXF2	1.47E-08
MA0613.1	FOXG1	2.13E-08
MA0067.1	Pax2	3.80E-08
MA0480.1	Foxo1	4.76E-08
MA0848.1	FOXO4	6.38E-08
MA0137.3	STAT1	8.47E-08
MA0849.1	FOXO6	8.87E-08
MA0633.1	Twist2	2.15E-07
MA0605.1	Atf3	3.26E-07
MA0025.1	NFIL3	6.05E-07
MA0028.2	ELK1	7.54E-07
MA0518.1	Stat4	7.91E-07
MA0504.1	NR2C2	9.63E-07
MA0786.1	POU3F1	1.21E-06
MA0910.1	Hoxd8	2.07E-06
MA0047.2	Foxa2	3.88E-06
MA0041.1	Foxd3	5.11E-06
MA0620.1	Mitf	6.73E-06
MA0601.1	Arid3b	8.61E-06
MA0042.2	FOX11	1.05E-05
MA0614.1	Foxj2	1.10E-05
MA0514.1	Sox3	1.25E-05
MA0507.1	POU2F2	1.93E-05
MA0788.1	POU3F3	2.33E-05
MA0790.1	POU4F1	3.65E-05
MA0913.1	Hoxd9	4.93E-05
MA0785.1	POU2F1	5.07E-05
MA0519.1	Stat5a::Stat5b	5.10E-05
MA0847.1	FOXO2	5.59E-05
MA0792.1	POU5F1B	6.65E-05
MA0791.1	POU4F3	8.11E-05
MA0692.1	TFEB	8.89E-05
MA0018.2	CREB1	9.62E-05
MA0829.1	Srebf1(var.2)	1.00E-04
MA0497.1	MEF2C	1.03E-04
MA0084.1	SRY	1.03E-04

MA0157.2	FOXO3	1.33E-04
MA0048.2	NHLH1	1.42E-04
MA0828.1	SREBF2(var.2)	1.50E-04
MA0687.1	SPIC	2.04E-04
MA0522.2	TCF3	2.46E-04
MA0831.1	TFE3	2.58E-04
MA0095.2	YY1	3.70E-04
MA0850.1	FOXP3	5.06E-04
MA0065.2	Pparg::Rxra	7.04E-04
MA0715.1	PROP1	8.06E-04
MA0784.1	POU1F1	8.15E-04
MA0787.1	POU3F2	9.71E-04
MA0809.1	TEAD4	1.00E-03
MA0871.1	TFEC	1.04E-03
MA0773.1	MEF2D	1.13E-03
MA0135.1	Lhx3	1.31E-03
MA0052.3	MEF2A	1.34E-03
MA0482.1	Gata4	1.59E-03
MA0703.1	LMX1B	1.70E-03
MA0660.1	MEF2B	2.43E-03
MA0520.1	Stat6	2.45E-03
MA0108.2	TBP	2.53E-03
MA0515.1	Sox6	2.60E-03
MA0830.1	TCF4	2.96E-03
MA0835.1	BATF3	3.09E-03
MA0789.1	POU3F4	5.92E-03
MA0683.1	POU4F2	6.31E-03
MA0909.1	HOXD13	7.71E-03
MA0056.1	MZF1	7.88E-03
MA0824.1	ID4	8.26E-03
MA0031.1	FOXD1	9.00E-03
MA0834.1	ATF7	9.07E-03
MA0093.2	USF1	1.39E-02
MA0485.1	Hoxc9	1.54E-02
MA0033.2	FOXL1	1.55E-02
MA0820.1	FIGLA	1.69E-02
MA0037.2	GATA3	2.01E-02
MA0899.1	HOXA10	2.39E-02
MA0772.1	IRF7	2.77E-02
MA0035.3	Gata1	3.93E-02
MA0139.1	CTCF	3.95E-02
MA0879.1	Dlx1	4.11E-02
MA0148.3	FOXA1	4.54E-02
MA0602.1	Arid5a	4.68E-02
MA0046.2	HNF1A	4.93E-02

8.5.2 Sites preferentially bound by HIF-2 in HepG2

Database ID	Transcription Factor	Adjusted <i>p</i> -value
MA0259.1	ARNT::HIF1A	1.95E-40
MA0847.1	FOXO2	6.60E-27
MA0033.2	FOXO1	1.05E-25
MA0846.1	FOXC2	1.19E-24
MA0047.2	Foxa2	1.60E-24
MA0850.1	FOXP3	5.31E-21
MA0148.3	FOXA1	1.04E-20
MA0032.2	FOXC1	4.61E-18
MA0851.1	Foxj3	6.44E-18
MA0042.2	FOXI1	7.46E-15
MA0845.1	FOXB1	7.46E-15
MA0848.1	FOXO4	3.12E-14
MA0481.1	FOXP1	8.62E-14
MA0523.1	TCF7L2	5.44E-13
MA0593.1	FOXP2	1.57E-12
MA0852.1	Foxk1	2.74E-12
MA0614.1	Foxj2	5.47E-12
MA0823.1	HEY1	1.33E-10
MA0849.1	FOXO6	2.02E-10
MA0031.1	FOXO1	4.96E-10
MA0004.1	Arnt	5.05E-10
MA0528.1	ZNF263	7.61E-10
MA0769.1	Tcf7	1.95E-09
MA0484.1	HNF4G	2.57E-09
MA0649.1	HEY2	4.75E-08
MA0768.1	LEF1	9.52E-08
MA0065.2	Pparg::Rxra	9.68E-08
MA0030.1	FOXF2	1.26E-06
MA1099.1	Hes1	2.04E-06
MA0160.1	NR4A2	3.13E-06
MA0041.1	Foxd3	3.46E-06
MA0613.1	FOXG1	8.44E-06
MA0480.1	Foxo1	2.60E-05
MA0677.1	Nr2f6	2.73E-05
MA0825.1	MNT	3.47E-05
MA0136.2	ELF5	7.78E-05
MA0081.1	SPIB	9.11E-05
MA0504.1	NR2C2	1.00E-04
MA0516.1	SP2	1.02E-04
MA0616.1	Hes2	1.13E-04
MA0617.1	Id2	2.55E-04
MA0098.3	ETS1	3.00E-04

MA0603.1	Arntl	4.23E-04
MA0079.3	SP1	4.42E-04
MA0017.2	NR2F1	4.53E-04
MA0002.2	RUNX1	5.26E-04
MA0076.2	ELK4	5.39E-04
MA0622.1	Mlxip	5.63E-04
MA0152.1	NFATC2	6.06E-04
MA0663.1	MLX	6.13E-04
MA0471.1	E2F6	6.82E-04
MA0514.1	Sox3	6.97E-04
MA0084.1	SRY	7.28E-04
MA0442.1	SOX10	7.59E-04
MA0464.2	BHLHE40	1.15E-03
MA0606.1	NFAT5	1.30E-03
MA0474.2	ERG	1.44E-03
MA0489.1	JUN(var.2)	2.09E-03
MA0505.1	Nr5a2	2.26E-03
MA0482.1	Gata4	2.56E-03
MA0515.1	Sox6	2.58E-03
MA0046.2	HNF1A	2.68E-03
MA0157.2	FOXO3	4.89E-03
MA0040.1	Foxq1	5.15E-03
MA0766.1	GATA5	7.34E-03
MA0162.2	EGR1	7.46E-03
MA0090.2	TEAD1	9.99E-03
MA0512.2	Rxra	1.14E-02
MA0599.1	KLF5	1.35E-02
MA0153.2	HNF1B	1.48E-02
MA0035.3	Gata1	1.61E-02
MA0645.1	ETV6	1.96E-02
MA0108.2	TBP	2.26E-02
MA0037.2	GATA3	2.35E-02
MA0476.1	FOS	2.39E-02
MA0490.1	JUNB	2.40E-02
MA0643.1	Esrrg	2.46E-02
MA0497.1	MEF2C	3.55E-02
MA0764.1	ETV4	3.73E-02
MA0636.1	BHLHE41	4.04E-02
MA0508.1	PRDM1	4.46E-02
MA0475.2	FLI1	4.84E-02
MA0156.2	FEV	4.99E-02
MA0762.1	ETV2	4.99E-02

8.5.3 Sites preferentially bound by HIF-1 in RCC4

Database ID	Transcription Factor	Adjusted p-value
MA0259.1	ARNT::HIF1A	2.22E-45
MA0528.1	ZNF263	2.85E-17
MA0649.1	HEY2	3.94E-17
MA0823.1	HEY1	1.58E-14
MA0004.1	Arnt	3.99E-14
MA0079.3	SP1	7.84E-14
MA0516.1	SP2	2.36E-13
MA0616.1	Hes2	1.61E-12
MA0599.1	KLF5	2.11E-12
MA0626.1	Npas2	1.34E-10
MA0663.1	MLX	4.47E-10
MA0136.2	ELF5	2.04E-09
MA0039.2	Klf4	7.25E-09
MA1099.1	Hes1	1.01E-08
MA0162.2	EGR1	1.31E-08
MA0076.2	ELK4	1.37E-08
MA0617.1	Id2	5.91E-08
MA0622.1	Mlxip	6.18E-08
MA0862.1	GMEB2	7.95E-08
MA0645.1	ETV6	8.44E-08
MA0471.1	E2F6	3.25E-07
MA0819.1	CLOCK	4.38E-07
MA0746.1	SP3	6.27E-07
MA0464.2	BHLHE40	1.04E-06
MA0603.1	Arntl	1.71E-06
MA0152.1	NFATC2	2.21E-06
MA0636.1	BHLHE41	2.88E-06
MA0747.1	SP8	6.02E-06
MA0493.1	Klf1	6.67E-06
MA0062.2	Gabpa	1.29E-05
MA0098.3	ETS1	1.63E-05
MA0825.1	MNT	1.68E-05
MA0474.2	ERG	3.13E-05
MA0762.1	ETV2	3.34E-05
MA0740.1	KLF14	8.06E-05
MA0058.3	MAX	9.02E-05
MA0685.1	SP4	1.46E-04
MA0489.1	JUN(var.2)	1.66E-04
MA0475.2	FLI1	2.45E-04
MA0741.1	KLF16	2.99E-04
MA0059.1	MAX::MYC	3.81E-04
MA0481.1	FOXP1	4.19E-04
MA0104.3	Mycn	5.57E-04
MA0598.2	EHF	6.56E-04
MA0490.1	JUNB	8.49E-04
MA0081.1	SPIB	8.53E-04
MA0655.1	JDP2	9.00E-04
MA0624.1	NFATC1	1.04E-03
MA0156.2	FEV	1.10E-03

MA0822.1	HES7	1.61E-03
MA0491.1	JUND	1.67E-03
MA0476.1	FOS	1.81E-03
MA0477.1	FOSL1	2.47E-03
MA0608.1	Creb3l2	2.67E-03
MA0478.1	FOSL2	2.71E-03
MA0664.1	MLXIPL	2.73E-03
MA0821.1	HES5	3.08E-03
MA0473.2	ELF1	3.15E-03
MA0625.1	NFATC3	3.48E-03
MA0641.1	ELF4	3.77E-03
MA0640.1	ELF3	4.46E-03
MA0764.1	ETV4	4.51E-03
MA0760.1	ERF	5.44E-03
MA0006.1	Ahr::Arnt	6.08E-03
MA0147.2	Myc	7.04E-03
MA0761.1	ETV1	7.07E-03
MA0733.1	EGR4	7.78E-03
MA0759.1	ELK3	9.49E-03
MA0606.1	NFAT5	1.02E-02
MA0472.2	EGR2	1.30E-02
MA0851.1	Foxj3	1.94E-02
MA0050.2	IRF1	2.17E-02
MA0593.1	FOXP2	2.61E-02
MA0852.1	Foxk1	2.75E-02
MA0514.1	Sox3	2.91E-02
MA0099.2	FOS::JUN	4.42E-02
MA0742.1	Klf12	5.13E-02
MA0841.1	NFE2	5.32E-02
MA0591.1	Bach1::Mafk	6.17E-02
MA0501.1	MAF::NFE2	6.72E-02
MA0517.1	STAT1::STAT2	6.81E-02
MA0604.1	Atf1	7.77E-02
MA0150.2	Nfe2l2	9.40E-02
MA0638.1	CREB3	1.03E-01
MA0686.1	SPDEF	1.06E-01
MA0067.1	Pax2	1.15E-01
MA0632.1	Tcf5	1.40E-01
MA0508.1	PRDM1	1.57E-01
MA0047.2	Foxa2	1.65E-01
MA0846.1	FOXC2	1.65E-01
MA0765.1	ETV5	1.97E-01
MA0615.1	Gmeb1	2.10E-01
MA0620.1	Mitf	2.11E-01
MA0597.1	THAP1	2.90E-01
MA0137.3	STAT1	3.08E-01
MA0028.2	ELK1	3.15E-01
MA0090.2	TEAD1	3.22E-01
MA0601.1	Arid3b	3.96E-01
MA0144.2	STAT3	4.81E-01

8.5.4 Sites preferentially bound by HIF-2 in RCC4

Database ID	Transcription Factor	Adjusted <i>p</i> -value
MA0259.1	ARNT::HIF1A	8.31E-26
MA0489.1	JUN(var.2)	1.18E-08
MA0476.1	FOS	6.29E-08
MA0490.1	JUNB	1.33E-07
MA0477.1	FOSL1	1.25E-06
MA0478.1	FOSL2	1.85E-06
MA0491.1	JUND	2.03E-06
MA0090.2	TEAD1	3.57E-06
MA0655.1	JDP2	3.83E-06
MA0823.1	HEY1	6.26E-06
MA0809.1	TEAD4	2.09E-05
MA1099.1	Hes1	3.81E-05
MA0099.2	FOS::JUN	1.27E-04
MA0462.1	BATF::JUN	4.33E-04
MA0649.1	HEY2	4.81E-04
MA0004.1	Arnt	4.87E-04
MA0002.2	RUNX1	5.41E-04
MA0136.2	ELF5	1.13E-03
MA0076.2	ELK4	2.45E-03
MA0528.1	ZNF263	2.47E-03
MA0825.1	MNT	5.16E-03
MA0808.1	TEAD3	8.17E-03
MA0841.1	NFE2	8.43E-03
MA0006.1	Ahr::Arnt	1.25E-02
MA0144.2	STAT3	2.04E-02
MA0591.1	Bach1::Mafk	2.61E-02
MA0081.1	SPIB	2.67E-02
MA0501.1	MAF::NFE2	2.96E-02
MA0645.1	ETV6	4.19E-02
MA0098.3	ETS1	4.28E-02

8.5.5 Sites preferentially bound by HIF-1 in HKC8

Database ID	Transcription Factor	Adjusted <i>p</i> -value
MA0259.1	ARNT::HIF1A	2.13E-39
MA0649.1	HEY2	6.58E-15
MA0528.1	ZNF263	1.12E-13
MA0004.1	Arnt	1.26E-13
MA0599.1	KLF5	2.27E-12
MA0823.1	HEY1	4.04E-12
MA0079.3	SP1	4.22E-12
MA0516.1	SP2	1.09E-11
MA0039.2	Klf4	1.61E-09
MA0136.2	ELF5	1.93E-09
MA0162.2	EGR1	4.70E-09
MA0076.2	ELK4	1.60E-08
MA1099.1	Hes1	1.65E-08
MA0152.1	NFATC2	2.17E-08
MA0645.1	ETV6	2.29E-08
MA0746.1	SP3	2.65E-08
MA0062.2	Gabpa	1.87E-07
MA0663.1	MLX	3.88E-07
MA0625.1	NFATC3	3.89E-07
MA0862.1	GMEB2	1.02E-06
MA0624.1	NFATC1	1.07E-06
MA0493.1	Klf1	1.71E-06
MA0067.1	Pax2	3.96E-06
MA0472.2	EGR2	5.66E-06
MA0617.1	Id2	5.84E-06
MA0598.2	EHF	5.93E-06
MA0641.1	ELF4	1.01E-05
MA0606.1	NFAT5	1.20E-05
MA0622.1	Mxiip	1.52E-05
MA0473.2	ELF1	1.85E-05
MA0050.2	IRF1	1.90E-05
MA0640.1	ELF3	2.41E-05
MA0747.1	SP8	3.81E-05
MA0464.2	BHLHE40	1.51E-04
MA0741.1	KLF16	1.63E-04
MA0616.1	Hes2	2.18E-04
MA0517.1	STAT1::STAT2	2.42E-04
MA0740.1	KLF14	2.76E-04
MA0474.2	ERG	2.81E-04
MA0471.1	E2F6	3.13E-04
MA0149.1	EWSR1-FLI1	3.16E-04
MA0603.1	Arntl	3.60E-04
MA0685.1	SP4	3.79E-04
MA0733.1	EGR4	4.29E-04
MA0475.2	FLI1	7.65E-04
MA0508.1	PRDM1	7.65E-04
MA0156.2	FEV	8.20E-04
MA0762.1	ETV2	8.95E-04
MA0058.3	MAX	9.19E-04
MA0098.3	ETS1	9.37E-04
MA0636.1	BHLHE41	9.98E-04
MA0518.1	Stat4	1.29E-03
MA0615.1	Gmeb1	1.33E-03
MA0137.3	STAT1	1.91E-03
MA0761.1	ETV1	2.06E-03
MA0759.1	ELK3	2.19E-03
MA0764.1	ETV4	2.48E-03
MA0144.2	STAT3	2.52E-03
MA0742.1	Klf12	3.15E-03
MA0732.1	EGR3	3.15E-03
MA0481.1	FOXP1	4.16E-03
MA0506.1	NRF1	4.40E-03
MA0593.1	FOXP2	1.20E-02
MA0664.1	MLXIPL	1.36E-02
MA0059.1	MAX::MYC	2.62E-02
MA0825.1	MNT	2.73E-02
MA0104.3	Mycn	2.90E-02
MA0028.2	ELK1	3.42E-02
MA0765.1	ETV5	3.59E-02
MA0006.1	Ahr::Arnt	4.06E-02
MA0822.1	HES7	4.68E-02
MA0626.1	Npas2	4.94E-02

8.5.6 Sites preferentially bound by HIF-2 in HKC8

Database ID	Transcription Factor	Adjusted <i>p</i> -value
MA0259.1	ARNT::HIF1A	1.09E-196
MA0489.1	JUN(var.2)	1.34E-118
MA0490.1	JUNB	9.77E-116
MA0491.1	JUND	4.05E-113
MA0478.1	FOSL2	7.53E-112
MA0476.1	FOS	1.76E-111
MA0477.1	FOSL1	2.70E-99
MA0809.1	TEAD4	2.03E-87
MA0099.2	FOS::JUN	2.67E-86
MA0090.2	TEAD1	2.94E-86
MA0655.1	JDP2	2.45E-84
MA0808.1	TEAD3	5.57E-76
MA0462.1	BATF::JUN	2.85E-71
MA1099.1	Hes1	2.85E-68
MA0004.1	Arnt	7.53E-60
MA0528.1	ZNF263	4.80E-56
MA0649.1	HEY2	2.08E-51
MA0823.1	HEY1	1.31E-49
MA0616.1	Hes2	1.86E-47
MA0841.1	NFE2	4.22E-47
MA0067.1	Pax2	4.44E-43
MA0825.1	MNT	2.87E-36
MA0481.1	FOXP1	1.11E-35
MA0136.2	ELF5	1.91E-35
MA0593.1	FOXP2	3.13E-35
MA0150.2	Nfe2l2	4.38E-35
MA0501.1	MAF::NFE2	2.05E-33
MA0081.1	SPIB	4.47E-33
MA0152.1	NFATC2	2.94E-32
MA0603.1	Arntl	2.95E-32
MA0847.1	FOXO2	8.27E-32
MA0079.3	SP1	3.53E-31
MA0516.1	SP2	6.47E-30
MA0850.1	FOXP3	3.37E-28
MA0599.1	KLF5	5.85E-28
MA0663.1	MLX	1.68E-27
MA0846.1	FOXC2	2.48E-27
MA0606.1	NFAT5	1.02E-26
MA0006.1	Ahr::Arnt	2.05E-26
MA0624.1	NFATC1	9.56E-26
MA0032.2	FOXC1	1.39E-25
MA0848.1	FOXO4	6.20E-25
MA0625.1	NFATC3	8.64E-25
MA0591.1	Bach1::Mafk	4.41E-24
MA0464.2	BHLHE40	8.89E-24
MA0162.2	EGR1	9.18E-24
MA0002.2	RUNX1	1.40E-23
MA0617.1	Id2	2.02E-23
MA0819.1	CLOCK	2.02E-23

MA0632.1	Tcf15	2.95E-23
MA0626.1	Npas2	8.54E-23
MA0622.1	Mlxip	1.19E-22
MA0845.1	FOXB1	2.70E-22
MA0862.1	GMEB2	1.92E-20
MA0042.2	FOXl1	2.88E-20
MA0033.2	FOXl1	3.16E-20
MA0144.2	STAT3	4.36E-20
MA0851.1	Foxj3	6.65E-20
MA0076.2	ELK4	7.89E-20
MA0089.1	MAFG::NFE2L1	8.49E-20
MA0849.1	FOXO6	1.38E-19
MA0031.1	FOXO1	1.67E-19
MA0852.1	Foxk1	2.39E-18
MA0746.1	SP3	3.32E-18
MA0604.1	Atf1	4.12E-18
MA0030.1	FOXF2	5.60E-18
MA0613.1	FOXG1	9.28E-18
MA0047.2	Foxa2	1.22E-17
MA0014.2	PAX5	1.47E-17
MA0137.3	STAT1	1.55E-17
MA0058.3	MAX	3.79E-17
MA0059.1	MAX::MYC	8.08E-17
MA0149.1	EWSR1-FLI1	9.89E-16
MA0104.3	Mycn	1.55E-15
MA0645.1	ETV6	6.02E-15
MA0899.1	HOXA10	6.63E-15
MA0620.1	Mitf	8.29E-15
MA0844.1	XBP1	1.91E-14
MA0157.2	FOXO3	3.29E-14
MA0605.1	Atf3	3.53E-14
MA0601.1	Arid3b	6.13E-14
MA0518.1	Stat4	6.35E-14
MA0614.1	Foxj2	7.40E-14
MA0471.1	E2F6	2.12E-13
MA0636.1	BHLHE41	6.23E-13
MA0480.1	Foxo1	8.60E-13
MA0485.1	Hoxc9	1.30E-12
MA0062.2	Gabpa	1.32E-12
MA0638.1	CREB3	1.33E-12
MA0508.1	PRDM1	1.83E-12
MA0482.1	Gata4	3.05E-12
MA0790.1	POU4F1	4.71E-12
MA0041.1	Foxd3	1.09E-11
MA0148.3	FOXA1	1.20E-11
MA0791.1	POU4F3	1.33E-11
MA0039.2	Klf4	2.09E-11
MA0598.2	EHF	2.16E-11
MA0909.1	HOXD13	2.51E-11
MA0608.1	Creb3l2	3.47E-11
MA0664.1	MLXIPL	4.39E-11
MA0048.2	NHLH1	5.13E-11
MA0607.1	Bhlha15	7.96E-11
MA0913.1	Hoxd9	8.19E-11
MA0050.2	IRF1	1.64E-10

MA0108.2	TBP	2.14E-10
MA0098.3	ETS1	2.27E-10
MA0147.2	Myc	3.17E-10
MA0519.1	Stat5a::Stat5b	3.20E-10
MA0517.1	STAT1::STAT2	5.53E-10
MA0822.1	HES7	6.34E-10
MA0065.2	Pparg::Rxra	6.72E-10
MA0609.1	Crem	7.60E-10
MA0640.1	ELF3	1.03E-09
MA0910.1	Hoxd8	1.25E-09
MA0497.1	MEF2C	1.36E-09
MA0017.2	NR2F1	3.03E-09
MA0499.1	Myod1	5.70E-09
MA0018.2	CREB1	8.01E-09
MA0046.2	HNF1A	1.17E-08
MA0037.2	GATA3	1.60E-08
MA0821.1	HES5	1.68E-08
MA0676.1	Nr2e1	2.47E-08
MA0500.1	Myog	2.79E-08
MA0878.1	CDX1	3.20E-08
MA0472.2	EGR2	6.14E-08
MA0905.1	HOXC10	7.54E-08
MA0816.1	Ascl2	1.20E-07
MA0474.2	ERG	1.20E-07
MA0901.1	HOXB13	1.86E-07
MA0035.3	Gata1	1.92E-07
MA0747.1	SP8	3.87E-07
MA0524.2	TFAP2C	3.91E-07
MA0762.1	ETV2	4.45E-07
MA0153.2	HNF1B	4.94E-07
MA0463.1	Bcl6	5.72E-07
MA0831.1	TFE3	7.10E-07
MA0597.1	THAP1	7.44E-07
MA0839.1	CREB3L1	9.02E-07
MA0135.1	Lhx3	9.52E-07
MA0764.1	ETV4	1.26E-06
MA0733.1	EGR4	1.56E-06
MA0521.1	Tcf12	1.81E-06
MA0786.1	POU3F1	1.92E-06
MA0811.1	TFAP2B	1.93E-06
MA0703.1	LMX1B	2.51E-06
MA0040.1	Foxq1	2.81E-06
MA0473.2	ELF1	2.85E-06
MA0641.1	ELF4	3.39E-06
MA0687.1	SPIC	6.12E-06
MA0475.2	FLI1	6.50E-06
MA0615.1	Gmrb1	8.82E-06
MA0522.2	TCF3	1.05E-05
MA0160.1	NR4A2	1.17E-05
MA0594.1	Hoxa9	1.37E-05
MA0810.1	TFAP2A(var.2)	1.62E-05
MA0520.1	Stat6	1.78E-05

MA0156.2	FEV	1.91E-05
MA0101.1	REL	2.45E-05
MA0651.1	HOXC11	2.47E-05
MA0759.1	ELK3	2.80E-05
MA0493.1	Klf1	2.97E-05
MA0003.3	TFAP2A	4.36E-05
MA0465.1	CDX2	5.06E-05
MA0504.1	NR2C2	5.16E-05
MA0812.1	TFAP2B(var.2)	7.03E-05
MA0692.1	TFEB	7.11E-05
MA0628.1	POU6F1	1.08E-04
MA0650.1	HOXA13	1.24E-04
MA0907.1	HOXC13	1.56E-04
MA0514.1	Sox3	1.60E-04
MA0741.1	KLF16	1.85E-04
MA0036.2	GATA2	1.90E-04
MA0683.1	POU4F2	2.88E-04
MA0908.1	HOXD11	3.01E-04
MA0911.1	Hoxa11	3.47E-04
MA0814.1	TFAP2C(var.2)	3.64E-04
MA0071.1	RORA	4.28E-04
MA0258.2	ESR2	5.98E-04
MA0761.1	ETV1	6.06E-04
MA0788.1	POU3F3	7.97E-04
MA0830.1	TCF4	1.42E-03
MA0119.1	NFIC::TLX1	1.47E-03
MA0635.1	BARHL2	1.50E-03
MA0684.1	RUNX3	1.69E-03
MA0511.2	RUNX2	1.85E-03
MA0526.1	USF2	2.36E-03
MA0507.1	POU2F2	3.41E-03
MA0906.1	HOXC12	3.56E-03
MA0785.1	POU2F1	3.72E-03
MA0787.1	POU3F2	4.70E-03
MA0781.1	PAX9	4.91E-03
MA0488.1	JUN	5.28E-03
MA0633.1	Twist2	5.81E-03
MA0052.3	MEF2A	5.86E-03
MA0092.1	Hand1::Tcf3	7.08E-03
MA0025.1	NFIL3	8.48E-03
MA0496.1	MAFK	9.40E-03
MA0793.1	POU6F2	9.79E-03
MA0698.1	ZBTB18	9.94E-03
MA0513.1	SMAD2::SMAD3::SMAD4	1.07E-02
MA0069.1	Pax6	1.73E-02
MA0792.1	POU5F1B	2.02E-02
MA0146.2	Zfx	2.61E-02
MA0784.1	POU1F1	2.81E-02
MA0691.1	TFAP4	2.87E-02
MA0834.1	ATF7	2.95E-02
MA0107.1	RELA	3.68E-02
MA0766.1	GATA5	4.86E-02

Chapter 9: Bibliography

- Acker, T., Diez-Juan, A., Aragonés, J., Tjwa, M., Brusselmans, K., Moons, L., Fukumura, D., Moreno-Murciano, M.P., Herbert, J.-M., Burger, A., et al. (2005). Genetic evidence for a tumor suppressor role of HIF-2 α . *Cancer Cell* 8, 131–141.
- Adhikary, S., and Eilers, M. (2005). Transcriptional regulation and transformation by Myc proteins. *Nat. Rev. Mol. Cell Biol.* 6, 635–645.
- Agalioti, T., Chen, G., and Thanos, D. (2002). Deciphering the transcriptional histone acetylation code for a human gene. *Cell* 111, 381–392.
- Alfranca, A., Gutiérrez, M.D., Vara, A., Aragonés, J., Vidal, F., Landázuri, M.O. (2002). c-Jun and hypoxia-inducible factor 1 functionally cooperate in hypoxia-induced gene transcription. *Molecular and cellular biology* 22, 12-22.
- Allen, J.W., Khetani, S.R., Johnson, R.S., and Bhatia, S.N. (2006). In vitro liver tissue model established from transgenic mice: role of HIF-1 α on hypoxic gene expression. *Tissue Eng* 12, 3135–3147.
- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166– 169.
- Bakker, W.J., Harris, I.S., and Mak, T.W. (2007). FOXO3a Is Activated in Response to Hypoxic Stress and Inhibits HIF1-Induced Apoptosis via Regulation of CITED2. *Mol. Cell* 28, 941–953.
- Benayoun, B.A., Caburet, S., and Veitia, R.A. (2011). Forkhead transcription factors: Key players in health and disease. *Trends Genet.* 27, 224–232.
- Biddie, S.C., John, S., Sabo, P.J., Thurman, R.E., Johnson, T.A., Schiltz, R.L., Miranda, T.B., Sung, M.-H., Trump, S., Lightman, S.L., et al. (2011). Transcription Factor AP1 Potentiates Chromatin Accessibility and Glucocorticoid Receptor Binding. *Mol Cell* 43, 145–155.
- Bieniossek, C., Papai, G., Schaffitzel, C., Garzoni, F., Chaillet, M., Scheer, E., Papadopoulos, P., Tora, L., Schultz, P., and Berger, I. (2013). The architecture of human general transcription factor TFIID core complex. *Nature* 493, 699–702.
- Bostrom, P., Magnusson, B., Svensson, P.A., Wiklund, O., Boren, J., Carlsson, L.M., Stahlman, M., Olofsson, S.O., and Hultén, L.M. (2006). Hypoxia converts human macrophages into triglyceride-loaded foam cells. *Arter. Thromb Vasc Biol* 26, 1871–1876.
- Botusan, I.R., Sunkari, V.G., Savu, O., Catrina, A.I., Grünler, J., Lindberg, S., Pereira, T., Ylä-Herttuala, S., Poellinger, L., Brismar, K., et al. (2008). Stabilization of HIF-1 α is critical to improve wound healing in diabetic mice. *Proc. Natl. Acad. Sci.* 105, 19426–19431.
- Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., and Crawford, G.E. (2008). High-resolution mapping and characterization of open chromatin across the genome. *Cell* 132, 311–322.
- Brem, H., Tomic-Canic, M., Entero, H., Hanflik, A.M., Wang, V.M., Fallon, J.T., and Ehrlich, H.P. (2007). The synergism of age and db/db genotype impairs wound healing. *Exp. Gerontol.* 42, 523–531.
- Brown, P.H., Alani, R., Preis, L.H., Szabo, E., and Birrer, M.J. (1993). Suppression of oncogene-induced transformation by a deletion mutant of c-jun. *Oncogene* 8, 877–886.
- Bruick, R.K., and McKnight, S.L. (2001). A conserved family of prolyl-4-hydroxylases that modify HIF. *Science* (80-.). 294, 1337–1340.
- Bussemaker, H.J., Li, H., and Siggia, E.D. (2001). Regulatory element detection using correlation with expression. *Nat. Genet.* 27, 167–171.
- Cai, Z., Zhong, H., Bosch-Marce, M., Fox-Talbot, K., Wang, L., Wei, C., Trush, M.A., and Semenza, G.L. (2008). Complete loss of ischaemic preconditioning-induced cardioprotection in mice with partial deficiency of HIF-1 α . *Cardiovasc. Res.* 77,

- 463–470.
- Carmeliet, P., Dor, Y., Herbert, J.M., Fukumura, D., Brusselmans, K., Dewerchin, M., Neeman, M., Bono, F., Abramovitch, R., Maxwell, P., et al. (1998). Role of HIF-1 α in hypoxia-mediated apoptosis, cell proliferation and tumour angiogenesis. *Nature* 394, 485–490.
- Carroll, J.S., Liu, X.S., Brodsky, A.S., Li, W., Meyer, C.A., Szary, A.J., Eeckhoute, J., Shao, W., Hestermann, E. V, Geistlinger, T.R., et al. (2005). Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* 122, 33–43.
- Chang, L., and Karin, M. (2001). Mammalian MAP kinase signalling cascades. *Nature* 410, 37–40.
- Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 14, 128.
- Chen, W., Hill, H., Christie, A., Kim, M.S., Holloman, E., Pavia-Jimenez, A., Homayoun, F., Ma, Y., Pater, N., Yell, P., Hao, G., Yousuf, Q., Joyce, A., Pedrosa, I., Geiger, H., Zhang, H., Chang, J., Gardner, K.H., Bruick, R.K., Reeves, C., Hwang, T.H., Courtney, K., Frenkel, E., Sun, X., Zojwalla, N., Wong, T., Rizzi, J.P., Wallace, E.M., Josey, J.A., Xie, Y., Xie, X.J., Kapur, P., McKay, R.M., Brugarolas, J. (2016). Targeting renal cell carcinoma with a HIF-2 antagonist. *Nature*. 539, 112–117.
- Chinenov, Y., and Kerppola, T.K. (2001). Close encounters of many kinds: Fos-Jun interactions that mediate transcription regulatory specificity. *Oncogene* 20, 2438–2452.
- Cho, H., Du, X., Rizzi, J.P., Liberzon, E., Chakroborty, A.A., Goa, W., Carvo, I.A., Signoretti, S., Bruick, R.K., Josey, J.A., Wallace, E.M., Kaelin Jr, W.G. (2016). On-Target Efficacy of a HIF2 α Antagonist in Preclinical Kidney Cancer Models. *Nature* 539, 107–111.
- Choudhry, H., Schodel, J., Oikonomopoulos, S., Camps, C., Grampp, S., Harris, A.L., Ratcliffe, P.J., Ragoussis, J., and Mole, D.R. (2014). Extensive regulation of the non-coding transcriptome by hypoxia: role of HIF in releasing paused RNAPol2. *EMBO Rep* 15, 70–76.
- Compernelle, V., Brusselmans, K., Acker, T., Hoet, P., Tjwa, M., Beck, H., Plaisance, S., Dor, Y., Keshet, E., Lupu, F., et al. (2002). Loss of HIF-2 α and inhibition of VEGF impair fetal lung maturation, whereas treatment with VEGF prevents fatal respiratory distress in premature mice. *Nat Med* 8, 702–710.
- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex Genome Engineering Using CRISPR/Cas System. *Science* (80-.). 339, 819–824.
- Corn, P.G., Ricci, M.S., Scata, K.A., Arsham, A.M., Simon, M.C., Dicker, D.T., and El-Deiry, W.S. (2005). Mxi1 is induced by hypoxia in a HIF-1- dependent manner and protects cells from c-Myc-induced apoptosis. *Cancer Biol Ther* 4, 1285–1294.
- Corominas, R., Yang, X., Lin, GN., Kang, S., Shen, Y., Ghamsari, L., Broly, M., Rodriguez, M., Tam, S., Trigg, S.A., Fan, C., Yi, S., Tasan, M., Lemmens, I., Kuang, X., Zhao, N., Malhotra, D., Michaelson, J.J., Vacic, V., Calderwood, M.A., Roth, F.P., Tavernier, J., Horvath, S., Salehi-Ashtiani, K., Korkin, D., Sebat, J., Hill, D.E., Hao, T., Vidal, M., and Iakoucheva, L.M. (2014). Protein interaction network of alternatively spliced isoforms from brain links genetic risk factors for autism. *Nat. Commun.* 5.
- Covello, K.L., Simon, M.C., and Keith, B. Targeted replacement of hypoxia-inducible factor-1 α by a hypoxia-inducible factor-2 α knock-in allele promotes tumor growth. (2005). *Cancer Res* 65, 2277–2286.
- Covello, K.L., Kehler, J., Yu, H., Gordan, J.D., Arsham, A.M., Hu, C.J., Labosky, P.A., Simon, M.C., and Keith, B. (2006). HIF-2 α regulates Oct-4: effects of hypoxia

- on stem cell function, embryonic development, and tumor growth. *Genes Dev* 20, 557–570.
- Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M. a, Frampton, G.M., Sharp, P. a, et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* 107, 21931–21936.
- Dale, R.K., Pedersen, B.S., and Quinlan, A.R. (2011). Pybedtools: A flexible Python library for manipulating genomic datasets and annotations. *Bioinformatics* 27, 3423–3424.
- Damert, A., Ikeda, E., and Risau, W. (1997). Activator-protein-1 binding potentiates the hypoxia-inducible factor-1-mediated hypoxia-induced transcriptional activation of vascular-endothelial growth factor expression in C6 glioma cells. *Biochem J* 327 (Pt 2), 419–423.
- Dang, C. V, Kim, J.W., Gao, P., and Yustein, J. (2008). The interplay between MYC and HIF in cancer. *Nat Rev Cancer* 8, 51–56.
- Davies, J.O.J., Telenius, J.M., McGowan, S.J., Roberts, N.A., Taylor, S., Higgs, D.R., and Hughes, J.R. (2015). Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nat. Methods* 13, 74–80.
- Dekker, J., Marti-Renom, M.A., and Mirny, L.A. (2013). Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat. Rev. Genet.* 14, 390–403.
- Desbaillets, I., Diserens, a C., de Tribolet, N., Hamou, M.F., and Van Meir, E.G. (1999). Regulation of interleukin-8 expression by reduced oxygen pressure in human glioblastoma. *Oncogene* 18, 1447–1456.
- Doedens, A.L., Stockmann, C., Rubinstein, M.P., Liao, D., Zhang, N., DeNardo, D.G., Coussens, L.M., Karin, M., Goldrath, A.W., and Johnson, R.S. (2010). Macrophage expression of hypoxia-inducible factor-1 alpha suppresses T-cell function and promotes tumor progression. *Cancer Res.* 70, 7465–7475.
- Dunwoodie, S.L. (2009). The Role of Hypoxia in Development of the Mammalian Embryo. *Dev. Cell* 17, 755–773.
- Eferl, R., and Wagner, E.F. (2003). AP-1: A double-edged sword in tumorigenesis. *Nat. Rev. Cancer* 3, 859–868.
- Elvidge, G.P., Glenny, L., Appelhoff, R.J., Ratcliffe, P.J., Ragoussis, J., and Gleadle, J.M. (2006). Concordant regulation of gene expression by hypoxia and 2-oxoglutarate-dependent dioxygenase inhibition: the role of HIF-1alpha, HIF-2alpha, and other pathways. *J Biol Chem* 281, 15215–15226.
- Ema, M., Morita, M., Ikawa, S., Tanaka, M., Matsuda, Y., Gotoh, O., Saijoh, Y., Fujii, H., Hamada, H., Kikuchi, Y., et al. (1996). Two new members of the murine Sim gene family are transcriptional repressors and show different expression patterns during mouse embryogenesis. *Mol. Cell. Biol.* 16, 5865–5875.
- Ema, M., Taya, S., Yokotani, N., Sogawa, K., Matsuda, Y., and Fujii-Kuriyama, Y. (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, 4273–4278.
- Epstein, A.C.R., Gleadle, J.M., McNeill, L.A., Hewitson, K.S., O'Rourke, J., Mole, D.R., Mukherji, M., Metzen, E., Wilson, M.I., Dhanda, A., et al. (2001). C-elegans EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. *Cell* 107, 43–54.
- Farnham, P.J. (2009). Insights from genomic profiling of transcription factors. *Nat. Rev. Genet.* 10, 605–616.
- Firth, J.D., Ebert, B.L., Pugh, C.W., and Ratcliffe, P.J. (1994). Oxygen-regulated control elements in the phosphoglycerate kinase 1 and lactate dehydrogenase A genes: Similarities with the erythropoietin 3' enhancer. *Genetics* 91, 6496–6500.
- Flamme, I., Fröhlich, T., Von Reutern, M., Kappel, A., Damert, A., and Risau, W. (1997). HRF, a putative basic helix-loop-helix-PAS-domain transcription factor is closely

- related to hypoxia-inducible factor-1 α and developmentally expressed in blood vessels. *Mech. Dev.* 63, 51–60.
- Furlan-Magaril, M., Rincón-Arango, H., Recillas-Targa, F. (2009). Sequential chromatin immunoprecipitation protocol: ChIP-reChIP. *Methods in Molecular Biology.* 543, 253–266.
- Gerlinger, M., Horswell, S., Larkin, J., Rowan, A.J., Salm, M.P., Varela, I., Fisher, R., McGranahan, N., Matthews, N., Santos, C.R., et al. (2014). Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat Genet* 46, 225–+.
- Gertz, J., Savic, D., Varley, K.E., Partridge, E.C., Safi, A., Jain, P., Cooper, G.M., Reddy, T.E., Crawford, G.E., and Myers, R.M. (2013). Distinct properties of cell-type-specific and shared transcription factor binding sites. *Mol Cell* 52, 25–36.
- Giatromanolaki, A., Koukourakis, M.I., Sivridis, E., Turley, H., Wykoff, C.C., Gatter, K.C., and Harris, A.L. (2003). DEC1 (STRA13) protein expression relates to hypoxia-inducible factor 1- α and carbonic anhydrase-9 overexpression in non-small cell lung cancer. *J. Pathol.* 200, 222–228.
- Gilkes, D.M., Xiang, L., Lee, S.J., Chaturvedi, P., Hubbi, M.E., Wirtz, D., and Semenza, G.L. (2014). Hypoxia-inducible factors mediate coordinated RhoA-ROCK1 expression and signaling in breast cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* 111, E384–93.
- Gordan, J.D., Thompson, C.B., and Simon, M.C. (2007). HIF and c-Myc: sibling rivals for control of cancer cell metabolism and proliferation. *Cancer Cell* 12, 108–113.
- Gossen, M., Freundlieb, S., Bender, G., Muller, G., Hillen, W., and Bujard, H. (1995). Transcriptional activation by tetracyclines in mammalian cells. *Science* (80-). 268, 1766–1769.
- Grabmaier, K., A de Weijert, M.C., Verhaegh, G. W., Schalken, J. A., and Oosterwijk, E. (2004). Strict regulation of CAIX(G250/MN) by HIF-1 α in clear cell renal cell carcinoma. *Oncogene* 23, 5624–31.
- Grampp, S., Platt, J.L., Lauer, V., Salama, R., Kranz, F., Neumann, V.K., Wach, S., Stohr, C., Hartmann, A., Eckardt, K.U., et al. (2016). Genetic variation at the 8q24.21 renal cancer susceptibility locus affects HIF binding to a MYC enhancer. *Nat. Commun.* 7.
- Grontved, L., John, S., Baek, S., Liu, Y., Buckley, J.R., Vinson, C., Aguilera, G., and Hager, G.L. (2013). C/EBP maintains chromatin accessibility in liver and facilitates glucocorticoid receptor recruitment to steroid response elements. *Embo J.* 32, 1568–1583.
- Gross, D.S., and Garrard, W.T. (1988). Nuclease hypersensitive sites in chromatin. *Annu Rev Biochem* 57, 159–197.
- Gruber, M., Hu, C.J., Johnson, R.S., Brown, E.J., Keith, B., and Simon, M.C. (2007). Acute postnatal ablation of Hif-2 α results in anemia. *Proc Natl Acad Sci U S A.* 104, 2301–2306.
- Grzenda, A., Lomberg, G., Zhang, J.S., and Urrutia, R. (2009). Sin3: master scaffold and transcriptional corepressor. *Biochim Biophys Acta* 1789, 443–450.
- Gu, Y.Z., Moran, S.M., Hogenesch, J.B., Wartman, L., and Bradfield, C.A. (1998). Molecular characterization and chromosomal localization of a third α -class hypoxia inducible factor subunit, HIF3 α . *Gene Expr* 7, 205–213.
- Guimbellot, J.S., Erickson, S.W., Mehta, T., Wen, H., Page, G.P., Sorscher, E.J., and Hong, J.S. (2009). Correlation of microRNA levels during hypoxia with predicted target mRNAs through genome-wide microarray analysis. *BMC Med Genomics* 2, 15.
- Gustafsson, M.V., Zheng, X., Pereira, T., Gradin, K., Jin, S., Lundkvist, J., Ruas, J.L., Poellinger, L., Lendahl, U., Bondesson, M. (2005). Hypoxia requires notch signaling to maintain the undifferentiated cell state. *Dev. Cell.* 9, 617–628.
- Hanahan, D., and Weinberg, R.A. (2000). The hallmarks of cancer. *Cell* 100, 57–70.
- Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: The next generation.

- Cell 144, 646–674.
- Hartley, I., Elkhoury, F.F., Heon Shin, J., Xie, B., Gu, X., Gao, Y., Zhou, D., and Haddad, G.G. (2013). Long-Lasting Changes in DNA Methylation Following Short-Term Hypoxic Exposure in Primary Hippocampal Neuronal Cultures. *PLoS One* 8.
- Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 39, 311–318.
- Hendrich, B., and Bird, A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell. Biol.* 18, 6538–6547.
- Hess, J., Angel, P., and Schorpp-Kistner, M. (2004). AP-1 subunits: quarrel and harmony among siblings. *J. Cell Sci.* 117, 5965–5973.
- Hewitson, K.S., McNeill, L.A., Riordan, M. V, Tian, Y.M., Bullock, A.N., Welford, R.W., Elkins, J.M., Oldham, N.J., Bhattacharya, S., Gleadle, J.M., et al. (2002). Hypoxia-inducible factor (HIF) asparagine hydroxylase is identical to factor inhibiting HIF (FIH) and is related to the cupin structural family. *J Biol Chem* 277, 26351–26355.
- Hill, C.S., Wynne, J., and Treisman, R. (1994). Serum-regulated transcription by serum response factor (SRF): a novel role for the DNA binding domain. *Embo J* 13, 5421–5432.
- Hoek, K.S., Schlegel, N.C., Eichhoff, O.M., Widmer, D.S., Praetorius, C., Einarsson, S.O., Valgeirsdottir, S., Bergsteinsdottir, K., Schepsky, A., Dummer, R., et al. (2008). Novel MITF targets identified using a two- step DNA microarray strategy. *Pigment Cell Melanoma Res.* 21, 665– 676.
- Hogenesch, J.B., Chan, W.K., Jackiw, V.H., Brown, R.C., Gu, Y.Z., PrayGrant, M., Perdew, G.H., and Bradfield, C.A. (1997). Characterization of a subset of the basic-helix-loop-helix-PAS superfamily that interacts with components of the dioxin signaling pathway. *J. Biol. Chem.* 272, 8581–8593.
- Hogenesch, J.B., Gu, Y.Z., Jain, S., and Bradfield, C.A. (1998). The basic-helix-loop-helix-PAS orphan MOP3 forms transcriptionally active complexes with circadian and hypoxia factors. *Proc. Natl. Acad. Sci. U. S. A.* 95, 5474–5479.
- Hower, V., Evans, S.N., and Pachter, L. (2011). Shape-based peak identification for ChIP-Seq. *BMC Bioinformatics* 12, 15.
- Hu, C.-J., Wang, L.-Y., Chodosh, L.A., Keith, B., and Simon, M.C. (2003). Differential roles of hypoxia-inducible factor 1alpha (HIF-1alpha) and HIF-2alpha in hypoxic gene regulation. *Mol. Cell. Biol.* 23, 9361–9374.
- Hughes, J.R., Roberts, N., McGowan, S., Hay, D., Giannoulatou, E., Lynch, M., De Gobbi, M., Taylor, S., Gibbons, R., and Higgs, D.R. (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat. Genet.* 46, 205–212.
- Hummel, K.P., Dickie, M.M., and Coleman, D.L. (1966). Diabetes, a New Mutation in the Mouse. *Science* (80-.). 153, 1127–1128.
- Imbalzano, a N., Kwon, H., Green, M.R., and Kingston, R.E. (1994). Facilitated binding of TATA-binding protein to nucleosomal DNA. *Nature* 370, 481–485.
- Irigoyen, M., Anso, E., Martinez, E., Garayoa, M., Martinez-Irujo, J.J., and Rouzaut, A. (2007). Hypoxia alters the adhesive properties of lymphatic endothelial cells. A transcriptional and functional study. *Biochim Biophys Acta* 1773, 880–890.
- Iyer, N. V, Kotch, L.E., Agani, F., Leung, S.W., Laughner, E., Wenger, R.H., Gassmann, M., Gearhart, J.D., Lawler, A.M., Yu, A.Y., et al. (1998). Cellular and developmental control of O₂ homeostasis by hypoxia-inducible factor 1 alpha. *Genes Dev* 12, 149–162.
- Jeffery, I.B., Madden, S.F., McGettigan, P.A., Perrière, G., Culhane, A.C., and Higgins, D.G. (2007). Integrating transcription factor binding site information with gene expression datasets. *Bioinformatics* 23, 298– 305.

- Ji, Z.L., Donaldson, I.J., Liu, J.R., Hayes, A., Zeef, L.A.H., and Sharrocks, A.D. (2012). The Forkhead Transcription Factor FOXK2 Promotes AP-1-Mediated Transcriptional Regulation. *Mol Cell Biol* 32, 385–398.
- Ji, Z.L., Mohammed, H., Webber, A., Ridsdale, J., Han, N., Carroll, J.S., and Sharrocks, A.D. (2014). The forkhead transcription factor FOXK2 acts as a chromatin targeting factor for the BAP1-containing histone deubiquitinase complex. *Nucleic Acids Res.* 42, 6232–6242.
- Jiang, B.H., Rue, E., Wang, G.L., Roe, R., and Semenza, G.L. (1996). Dimerization, DNA binding, and transactivation properties of hypoxia-inducible factor 1. *J Biol Chem* 271, 17771–17778.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A Programmable Dual-RNA – Guided. *Science* (80-.). 337, 816–822.
- Jones, S. (2004). An overview of the basic helix-loop-helix proteins. *Genome Biol.* 5, 226.
- Kaelin, W.G. (2007). Von Hippel-Lindau disease. *Annu Rev Pathol* 2, 145–173.
- Kaelin, W.G., and Ratcliffe, P.J. (2008). Oxygen sensing by metazoans: The central role of the HIF hydroxylase pathway. *Mol Cell* 30, 393–402.
- Kanno, T., Kamba, T., Yamasaki, T., Shibasaki, N., Saito, R., Terada, N., Toda, Y., Mikami, Y., Inoue, T., Kanematsu, A., et al. (2012). JunB promotes cell invasion and angiogenesis in VHL-defective renal cell carcinoma. *Oncogene* 31, 3098–3110.
- Kasper, L.H., Boussouar, F., Boyd, K., Xu, W., Biesen, M., Rehg, J., Baudino, T.A., Cleveland, J.L., and Brindle, P.K. (2005). Two transactivation mechanisms cooperate for the bulk of HIF-1-responsive gene expression. *Embo J* 24, 3846–3858.
- Kato, Y., Kawamoto, T., Fujimoto, K., and Noshiro, M. (2014). DEC1/STRA13/SHARP2 and DEC2/SHARP1 Coordinate Physiological Processes, Including Circadian Rhythms in Response to Environmental Stimuli. In *Bhlh Transcription Factors in Development and Disease*, R. Taneja, ed. (San Diego: Elsevier Academic Press Inc), pp. 339–372.
- Kawata, H., Yamada, K., Shou, Z., Mizutani, T., Yazawa, T., Yoshino, M., Sekiguchi, T., Kajitani, T., and Miyamoto, K. (2003). Zinc-fingers and homeoboxes (ZHX) 2, a novel member of the ZHX family, functions as a transcriptional repressor. *Biochem. J.* 373, 747–757.
- Keith, B., Johnson, R.S., and Simon, M.C. (2012). HIF1alpha and HIF2alpha: sibling rivalry in hypoxic tumour growth and progression. *Nat Rev Cancer* 12, 9–22.
- Kewley, R.J., Whitelaw, M.L., and Chapman-Smith, A. (2004). The mammalian basic helix-loop-helix/PAS family of transcriptional regulators. *Int J Biochem Cell Biol* 36, 189–204.
- Kim, J.W., Tchernyshyov, I., Semenza, G.L., and Dang, C. V (2006). HIF-1-mediated expression of pyruvate dehydrogenase kinase: a metabolic switch required for cellular adaptation to hypoxia. *Cell Metab* 3, 177–185.
- Koh, M.Y., Darnay, B.G., and Powis, G. (2008). Hypoxia-associated factor, a novel E3-ubiquitin ligase, binds and ubiquitinates hypoxia-inducible factor 1alpha, leading to its oxygen-independent degradation. *Mol. Cell. Biol.* 28, 7081–7095.
- Koh, M.Y., Lemos, R., Liu, X., and Powis, G. (2011). The hypoxia-associated factor switches cells from HIF-1 α - to HIF-2 α -dependent signaling promoting stem cell characteristics, aggressive tumor growth and invasion. *Cancer Res.* 71, 4015–4027.
- Koizume, S., and Miyagi, Y. (2015). Diverse mechanisms of Sp1-Dependent transcriptional regulation potentially involved in the adaptive response of cancer cells to oxygen-deficient conditions. *Cancers (Basel).* 8.
- Laderoute, K.R. (2005). The interaction between HIF-1 and AP-1 transcription factors in response to low oxygen. *Semin. Cell Dev. Biol.* 16, 502–513.
- Laganière, J., Deblois, G., Lefebvre, C., Bataille, A.R., Robert, F., and Giguère, V.

- (2005). From the Cover: Location analysis of estrogen receptor alpha target promoters reveals that FOXA1 defines a domain of the estrogen response. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 11651–11656.
- Lam, E.W.-F., Brosens, J.J., Gomes, A.R., and Koo, C.-Y. (2013). Forkhead box proteins: tuning forks for transcriptional harmony. *Nat. Rev. Cancer* **13**, 482–495.
- Lando, D., Peet, D.J., Gorman, J.J., Whelan, D.A., Whitelaw, M.L., and Bruick, R.K. (2002). FIH-1 is an asparaginyl hydroxylase enzyme that regulates the transcriptional activity of hypoxia-inducible factor. *Genes Dev* **16**, 1466–1471.
- Lane, N., and Martin, W. (2010). The energetics of genome complexity. *Nature* **467**, 929–934.
- Lau, K.W., Tian, Y., Raval, R.R., Ratcliffe, P.J., Pugh, C.W. (2007). Target gene selectivity of hypoxia-inducible factor- α in renal cancer cells is conveyed by post-DNA-binding mechanisms. *Br J Cancer* **96**, 1284–1292.
- Ledaki, I., McIntyre, A., Wigfield, S., Buffa, F., McGowan, S., Baban, D., Li, J., and Harris, A.L. (2015). Carbonic anhydrase IX induction defines a heterogeneous cancer cell response to hypoxia and mediates stem cell-like properties and sensitivity to HDAC inhibition. *Oncotarget* **6**, 19413–19427.
- Lee, Y.J., and Corry, P.M. (1999). Hypoxia-induced bFGF gene expression is mediated through the JNK signal transduction pathway. *Mol Cell Biochem* **202**, 1–8.
- Lee, W., Mitchell, P., and Tjian, R. (1987). Purified transcription factor AP-1 interacts with TPA-inducible enhancer elements. *Cell* **49**, 741–752.
- Lees, M.J., and Whitelaw, M.L. (1999). Multiple roles of ligand in transforming the dioxin receptor to an active basic helix-loop-helix/PAS transcription factor complex with the nuclear protein Arnt. *Mol. Cell. Biol.* **19**, 5811–5822.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760.
- Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. *Cell* **128**, 707–719.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009a). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079.
- Li, X., He, Y., Xu, Y., Huang, X., Liu, J., Xie, M., and Liu, X. (2016). KLF5 mediates vascular remodeling via HIF-1 α in hypoxic pulmonary hypertension. *Am. J. Physiol. - Lung Cell. Mol. Physiol.* **310**, L299–L310.
- Li, Z., Bao, S., Wu, Q., Wang, H., Eyler, C., Sathornsumetee, S., Shi, Q., Cao, Y., Lathia, J., McLendon, R.E., et al. (2009b). Hypoxia-Inducible Factors Regulate Tumorigenic Capacity of Glioma Stem Cells. *Cancer Cell* **15**, 501–513.
- Liao, D., Corle, C., Seagroves, T.N., and Johnson, R.S. (2007). Hypoxia-inducible factor-1 α is a key regulator of metastasis in a transgenic model of cancer initiation and progression. *Cancer Res.* **67**, 563–572.
- Lidgren, A., Hedberg, Y., Grankvist, K., Rasmuson, T., Vasko, J., and Ljungberg, B. (2005). The expression of hypoxia-inducible factor 1 α is a favorable independent prognostic factor in renal cell carcinoma. *Clin. Cancer Res.* **11**, 1129–1135.
- Liu, X., Li, H., Rajurkar, M., Li, Q., Cotton, J.L., Ou, J., Zhu, L.J., Goel, H.L., Mercurio, A.M., Park, J.S., et al. (2016). Tead and AP1 Coordinate Transcription and Motility. *Cell Rep.* **14**, 1169–1180.
- Loenarz, C., Coleman, M.L., Boleininger, A., Schierwater, B., Holland, P.W., Ratcliffe, P.J., and Schofield, C.J. (2011). The hypoxia-inducible transcription factor pathway regulates oxygen sensing in the simplest animal, *Trichoplax adhaerens*. *EMBO Rep* **12**, 63–70.
- Long, H.K., Prescott, S.L., and Wysocka, J. (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* **167**, 1170–1187.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550.

- Mace, K.A., Yu, D.H., Paydar, K.Z., Boudreau, N., and Young, D.M. (2007). Sustained expression of Hif-1alpha in the diabetic environment promotes angiogenesis and cutaneous wound repair. *Wound Repair Regen.* **15**, 636–645.
- Mahon, P.C., Hirota, K., and Semenza, G.L. (2001). FIH-1: a novel protein that interacts with HIF-1alpha and VHL to mediate repression of HIF-1 transcriptional activity. *Genes Dev* **15**, 2675–2686.
- Maranchie, J.K., Vasselli, J.R., Riss, J., Bonifacino, J.S., Linehan, W.M., and Klausner, R.D. The contribution of VHL substrate binding and HIF1-alpha to the phenotype of VHL loss in renal cell carcinoma. (2002). *Cancer Cell* **1**, 247–255.
- Maxwell, P.H., Pugh, C.W., and Ratcliffe, P.J. (1993). Inducible operation of the erythropoietin 3' enhancer in multiple cell lines: evidence for a widespread oxygen-sensing mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 2423–2427.
- Maxwell, P.H., Wiesener, M.S., Chang, G.W., Clifford, S.C., Vaux, E.C., Cockman, M.E., Wykoff, C.C., Pugh, C.W., Maher, E.R., and Ratcliffe, P.J. (1999). The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature* **399**, 271–275.
- Maynard, M.A., Qi, H., Chung, J., Lee, E.H., Kondo, Y., Hara, S., Conaway, R.C., Conaway, J.W., and Ohh, M. (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**, 11032–11040.
- Maynard, M.A., Evans, A.J., Hosomi, T., Hara, S., Jewett, M.A., and Ohh, M. (2005). Human HIF-3alpha4 is a dominant-negative regulator of HIF-1 and is down-regulated in renal cell carcinoma. *Faseb J* **19**, 1396–406.
- Mazumdar, J., Hickey, M.M., Pant, D.K., Durham, A.C., Sweet-Cordero, A., Vachani, A., Jacks, T., Chodosh, L.A., Kissil, J.L., Simon, M.C., et al. (2010). HIF-2alpha deletion promotes Kras-driven lung tumor development. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 14182–14187.
- McLeay, R.C., and Bailey, T.L. (2010). Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinformatics* **11**, 165.
- McNeill, L.A., Hewitson, K.S., Gleadle, J.M., Horsfall, L.E., Oldham, N.J., Maxwell, P.H., Pugh, C.W., Ratcliffe, P.J., and Schofield, C.J. (2002). The use of dioxygen by HIF prolyl hydroxylase (PHD1). *Bioorg Med Chem Lett* **12**, 1547–1550.
- Menrad, H., Werno, C., Schmid, T., Copanaki, E., Deller, T., Dehne, N., and Brüne, B. (2010). Roles of hypoxia-inducible factor-1α (HIF-1α) versus HIF-2α in the survival of hepatocellular tumor spheroids. *Hepatology* **51**, 2183–2192.
- Mense, S.M., Sengupta, A., Zhou, M., Lan, C., Bentsman, G., Volsky, D.J., and Zhang, L. (2006). Gene expression profiling reveals the profound upregulation of hypoxia-responsive genes in primary human astrocytes. *Physiol Genomics* **25**, 435–449.
- Mohammed, H., D'Santos, C., Serandour, A.A., Ali, H.R., Brown, G.D., Atkins, A., Rueda, O.M., Holmes, K.A., Theodorou, V., Robinson, J.L., et al. (2013). Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. *Cell Rep* **3**, 342–349.
- Mole, D.R., Blancher, C., Copley, R.R., Pollard, P.J., Gleadle, J.M., Ragoussis, J., and Ratcliffe, P.J. (2009). Genome-wide association of hypoxia-inducible factor (HIF)-1alpha and HIF-2alpha DNA binding with expression profiling of hypoxia-inducible transcripts. *J Biol Chem* **284**, 16767–16775.
- Muñoz-Nájara, U.M., Neurath, K.M., Vumbaca, F., and Claffey, K.P. (2006). Hypoxia stimulates breast carcinoma cell invasion through MT1-MMP and MMP-2 activation. *Oncogene* **25**, 2379–2392.
- Musselman, C.A., Lalonde, M.-E., Côté, J., and Kutateladze, T.G. (2012). Perceiving the epigenetic landscape through histone readers. *Nat. Struct. Mol. Biol.* **19**, 1218–1227.
- Nakato, R., and Shirahige, K. (2016). Recent advances in ChIP-seq analysis: from quality management to whole-genome annotation. *Brief. Bioinform.* **bbw023**.

- Nishida, H., Suzuki, T., Kondo, S., Miura, H., Fujimura, Y.I., and Hayashizaki, Y. (2006). Histone H3 acetylated at lysine 9 in promoter is associated with low nucleosome density in the vicinity of transcription start site in human cell. *Chromosom. Res.* 14, 203–211.
- Ogryzko, V. V., Schiltz, R.L., Russanova, V., Howard, B.H., and Nakatani, Y. (1996). The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell* 87, 953–959.
- Olive, M., Krylov, D., Echlin, D.R., Gardner, K., Taparowsky, E., and Vinson, C. (1997). A dominant negative to activation protein-1 (AP1) that abolishes DNA binding and inhibits oncogenesis. *J. Biol. Chem.* 272, 18586–18594.
- Ortiz-Barahona, A., Villar, D., Pescador, N., Amigo, J., and del Peso, L. (2010). Genome-wide identification of hypoxia-inducible factor binding sites and target genes by a probabilistic model integrating transcription-profiling data and in silico binding site prediction. *Nucleic Acids Res* 38, 2332–2345.
- Ozanne, B.W., Spence, H.J., McGarry, L.C., and Hennigan, R.F. (2007). Transcription factors control invasion: AP-1 the first among equals. *Oncogene* 26, 1–10.
- Papandreou, I., Cairns, R.A., Fontana, L., Lim, A.L., and Denko, N.C. (2006). HIF-1 mediates adaptation to hypoxia by actively downregulating mitochondrial oxygen consumption. *Cell Metab* 3, 187–197.
- Park, P.J. (2009). ChIP-seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.* 10, 669–680.
- Park, S.K., Dadak, A.M., Haase, V.H., Fontana, L., Giaccia, A.J., and Johnson, R.S. (2003). Hypoxia-induced gene expression occurs solely through the action of hypoxia-inducible factor 1alpha (HIF-1alpha): role of cytoplasmic trapping of HIF-2alpha. *Mol Cell Biol* 23, 4959–4971.
- Peng, J., Zhang, L., Drysdale, L., and Fong, G.H. (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, 8386–8391.
- Perez-Perri, J.I., Dengler, V.L., Audetat, K.A., Pandey, A., Bonner, E.A., Urh, M., Mendez, J., Daniels, D.L., Wappner, P., Galbraith, M.D., et al. (2016). The TIP60 Complex Is a Conserved Coactivator of HIF1A. *Cell Rep.* 16, 37–47.
- Phatnani, H.P., and Greenleaf, A.L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.* 20, 2922–2936.
- Pietras, A., Hansford, L.M., Johnsson, A.S., Bridges, E., Sjolund, J., Gisselsson, D., Rehn, M., Beckman, S., Noguera, R., Navarro, S., et al. (2009). HIF-2alpha maintains an undifferentiated state in neural crest- like human neuroblastoma tumor-initiating cells. *Proc. Natl. Acad. Sci. U. S. A.* 106, 16805–16810.
- Platt, J.L., Salama, R., Smythies, J., Choudhry, H., Davies, J.O.J., Hughes, J.R., Ratcliffe, P.J., and Mole, D.R. (2016). Capture-C reveals preformed chromatin interactions between HIF-binding sites and distant promoters. *EMBO Rep* 17, 1410–1421.
- Probst, M.R., Fan, C.M., Tessier-Lavigne, M., and Hankinson, O. (1997). Two murine homologs of the Drosophila single-minded protein that interact with the mouse aryl hydrocarbon receptor nuclear translocator protein. *J. Biol. Chem.* 272, 4451–4457.
- Rankin, E.B., Higgins, D.F., Walisser, J.A., Johnson, R.S., Bradfield, C.A., and Haase, V.H. (2005). Inactivation of the arylhydrocarbon receptor nuclear translocator (Arnt) suppresses von Hippel-Lindau disease-associated vascular tumors in mice. *Mol. Cell Biol.* 25, 3163–3172.
- Raval, R.R., Lau, K.W., Tran, M.G., Sowter, H.M., Mandriota, S.J., Li, J.L., Pugh, C.W., Maxwell, P.H., Harris, A.L., and Ratcliffe, P.J. (2005). Contrasting properties of hypoxia-inducible factor 1 (HIF-1) and HIF-2 in von Hippel-Lindau-associated renal cell carcinoma. *Mol Cell Biol* 25, 5675–5686.
- Ravasi, T., Suzuki, H., Cannistraci, C.V., Katayama, S., Bajic, V.B., Tan, K., Akalin, A., Schmeier, S., Kanamori-Katayama, M., Bertin, N., Carninci, P., Daub, C.O.,

- Forrest, A.R., Gough, J., Grimmond, S., Han, J.H., Hashimoto, T., Hide, W., Hofmann, O., Kamburov, A., Kaur, M., Kawaji, H., Kubosaki, A., Lassmann, T., van Nimwegen, E., MacPherson, C.R., Ogawa, C., Radovanovic, A., Schwartz, A., Teasdale, R.D., Tegner, J., Lenhard, B., Teichmann, S.A., Arakawa, T., Ninomiya, N., Murakami, K., Tagami, M., Fukuda, S., Imamura, K., Kai, C., Ishihara, R., Kitazume, Y., Kawai, J., Hume, D.A., Ideker, T., and Hayashizaki, Y. (2010). An atlas of combinatorial transcriptional regulation in mouse and man. *Cell. Signal.* **140**, 744–752.
- Ray, J.B., Arab, S., Deng, Y., Liu, P., Penn, L., Courtman, D.W., and Ward, M.E. (2008). Oxygen regulation of arterial smooth muscle cell proliferation and survival. *Am J Physiol Hear. Circ Physiol* **294**, H839–52.
- Robinson, C.M., Neary, R., Leventale, A., Watson, C.J., and Baugh, J.A. (2012). Hypoxia-induced DNA hypermethylation in human pulmonary fibroblasts is associated with Thy-1 promoter methylation and the development of a pro-fibrotic phenotype. *Respir. Res.* **13**, 74.
- Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., et al. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389–393.
- Rotem, A., Ram, O., Shores, N., Sperling, R.A., Goren, A., Weitz, D.A., Bernstein, B.E. (2015). Single-cell ChIP-seq reveals subpopulations defined by chromatin state. *Nat. Biotechnol.* **33**, 1165–1172.
- Ryan, H.E., Lo, J., and Johnson, R.S. (1998). HIF-1 α is required for solid tumour formation and embryonic vascularization. *Embo J.* **17**, 3005–3015.
- Ryan, H.E., Poloni, M., McNulty, W., Elson, D., Gassmann, M., Arbeit, J.M., and Johnson, R.S. (2000). Hypoxia-inducible factor-1 α is a positive factor in solid tumor growth. *Cancer Res* **60**, 4010–4015.
- Rytkönen, K.T., Williams, T.A., Renshaw, G.M., Primmer, C.R., and Nikinmaa, M. (2011). Molecular evolution of the metazoan PHD-HIF oxygen-sensing system. *Mol. Biol. Evol.* **28**, 1913–1926.
- Salama, R., Masson, N., Simpson, P., Sciesielski, L.K., Sun, M., Tian, Y.-M., Ratcliffe, P.J., and Mole, D.R. (2015). Heterogeneous Effects of Direct Hypoxia Pathway Activation in Kidney Cancer. *PLoS One* **10**.
- Sander, J.D., and Joung, J.K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* **32**, 347–355.
- Saxonov, S., Berg, P., and Brutlag, D.L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 1412–1417.
- Schipani, E., Ryan, H.E., Didrickson, S., Kobayashi, T., Knight, M., and Johnson, R.S. (2001). Hypoxia in cartilage: HIF-1 α is essential for chondrocyte growth arrest and survival. *Genes Dev.* **15**, 2865–2876.
- Schodel, J., Oikonomopoulos, S., Ragoussis, J., Pugh, C.W., Ratcliffe, P.J., and Mole, D.R. (2011). High-resolution genome-wide mapping of HIF-binding sites by ChIP-seq. *Blood* **117**, e207–17.
- Schodel, J., Bardella, C., Sciesielski, L.K., Brown, J.M., Pugh, C.W., Buckle, V., Tomlinson, I.P., Ratcliffe, P.J., and Mole, D.R. (2012). Common genetic variants at the 11q13.3 renal cancer susceptibility locus influence binding of HIF to an enhancer of cyclin D1 expression. *Nat Genet* **44**, 420–425, S1–2.
- Schodel, J., Mole, D.R., and Ratcliffe, P.J. (2013). Pan-genomic binding of hypoxia-inducible transcription factors. *Biol Chem* **394**, 507–517.
- Schulz, K., Milke, L., Rübsamen, D., Menrad, H., Schmid, T., and Brüne, B. (2012). HIF-1 α protein is upregulated in HIF-2 α depleted cells via enhanced translation. *FEBS Lett.* **586**, 1652–1657.
- Scortegagna, M., Ding, K., Oktay, Y., Gaur, A., Thurmond, F., Yan, L.J., Marck, B. T., Matsumoto, A.M., Shelton, J.M., Richardson, J.A., Bennett, M.J., and Garcia, J.A.

- Multiple organ pathology, metabolic abnormalities and impaired homeostasis of reactive oxygen species in *Epas1*^{-/-} mice. (2003). *Nat Genet* 35, 331–340.
- Semenza, G.L. (2003). Targeting HIF-1 for cancer therapy. *Nat Rev Cancer* 3, 721–732.
- Semenza, G.L. (2012). Hypoxia-inducible factors in physiology and medicine. *Cell* 148, 399–408.
- Semenza, G., Jiang, B., Leung, S., Passantino, R., Concordet, J., Maire, P., and Giallongo, A. (1996). Hypoxia Response Elements in the Aldolase A, Enolase 1, and Lactate Dehydrogenase A Gene Promoters Contain Essential Binding Sites for Hypoxia-inducible Factor 1. *J. Biol. Chem.* 271, 32529–32537.
- Semenza, G.L., Neifelt, M.K., Chi, S.M., and Antonarakis, S.E. (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, 5680–5684.
- Semenza, G.L., Roth, P.H., Fang, H.M., and Wang, G.L. (1994). Transcriptional regulation of genes encoding glycolytic enzymes by hypoxia-inducible factor 1. *J Biol Chem* 269, 23757–23763.
- Sharma, K., McCue, P., and Dunn, S.R. (2003). Diabetic kidney disease in the db/db mouse. *Am. J. Physiol. Renal Physiol.* 284, F1138–F1144.
- Shaulian, E. (2010). AP-1-The Jun proteins: Oncogenes or tumor suppressors in disguise? *Cellular Signalling* 22, 894–899.
- Shen, L., Shao, N., Liu, X., and Nestler, E. (2014). ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* 15, 284.
- Shi, X., Hong, T., Walter, K.L., Ewalt, M., Michishita, E., Hung, T., Carney, D., Peña, P., Lan, F., Kaadige, M.R., et al. (2006). ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature*.
- Shinojima, T., Oya, M., Takayanagi, A., Mizuno, R., Shimizu, N., and Murai, M. (2007). Renal cancer cells lacking hypoxia inducible factor (HIF)-1 α expression maintain vascular endothelial growth factor expression through HIF-2 α . *Carcinogenesis* 28, 529–536.
- Shohet, R. V., and Garcia, J.A. (2007). Keeping the engine primed: HIF factors as key regulators of cardiac metabolism and angiogenesis during ischemia. *J. Mol. Med.* 85, 1309–1315.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., et al. (2014). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539–539.
- Solomon, M.J., and Varshavsky, A. (1985). Formaldehyde-mediated DNA-protein crosslinking: A probe for in vivo chromatin structures (simian virus 40 chromosomes/nucleosome-free region/Pronase-resistant crosslinks/lac repressor/a-protein). *Biochemistry* 63, 6130–6134.
- Sowter, H.M., Raval, R.R., Moore, J.W., Ratcliffe, P.J., Harris, A.L. (2003). Predominant role of hypoxia-inducible transcription factor (Hif)-1 α versus Hif-2 α in regulation of the transcriptional response to hypoxia. *Cancer Res.* 65, 3299–3306.
- Steunou, A.L., Ducoux-Petit, M., Lazar, I., Monsarrat, B., Erard, M., Muller, C., Clottes, E., Burlet-Schiltz, O., and Nieto, L. (2013). Identification of the hypoxia-inducible factor 2 α nuclear interactome in melanoma cells reveals master proteins involved in melanoma development. *Mol Cell Proteomics* 12, 736–748.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102, 15545–15550.
- Takubo, K., Goda, N., Yamada, W., Iriuchishima, H., Ikeda, E., Kubota, Y., Shima, H., Johnson, R.S., Hirao, A., Suematsu, M., et al. (2010). Regulation of the HIF-

- 1alpha level is essential for hematopoietic stem cells-SOM. *Cell Stem Cell* 7, 391–402.
- Talks, K.L., Turley, H., Gatter, K.C., Maxwell, P.H., Pugh, C.W., Ratcliffe, P.J., and Harris, A.L. (2000). The expression and distribution of the hypoxia-inducible factors HIF-1alpha and HIF-2alpha in normal human tissues, cancers, and tumor-associated macrophages. *Am J Pathol* 157, 411–421.
- Tanaka, T., Wiesener, M.S., Bernhardt, W., Eckardt, K.U., and Warnecke, C. (2009). The human hypoxia-inducible factor (HIF)-3alpha gene is a HIF-1 target and may modulate hypoxic gene induction. *Biochem J*.
- Tanimoto, K., Tsuchihara, K., Kanai, A., Arauchi, T., Esumi, H., Suzuki, Y., and Sugano, S. (2010). Genome-wide identification and annotation of HIF-1alpha binding sites in two cell lines using massively parallel sequencing. *Hugo J* 4, 35–48.
- Tausendschön, M., Rehli, M., Dehne, N., Schmidl, C., Döring, C., Hansmann, M.L., and Brüne, B. (2015). Genome-wide identification of hypoxia-inducible factor-1 and -2 binding sites in hypoxic human macrophages alternatively activated by IL-10. *Biochim. Biophys. Acta - Gene Regul. Mech.* 1849, 10–22.
- Thienpont, B., Steinbacher, J., Zhao, H., D'Anna, F., Kuchnio, A., Ploumakis, A., Ghesquière, B., Van Dyck, L., Boeckx, B., Schoonjans, L., et al. (2016). Tumour hypoxia causes DNA hypermethylation by reducing TET activity. *Nature* 537, 63–68.
- Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernet, B., et al. (2012). The accessible chromatin landscape of the human genome. *Nature* 489, 75–82.
- Tian, H., McKnight, S.L., and Russell, D.W. (1997). Endothelial PAS domain protein 1 (EPAS1), a transcription factor selectively expressed in endothelial cells. *Genes Dev* 11, 72–82.
- Tian, H., Hammer, R.E., Matsumoto, A.M., Russell, D.W., and McKnight, S.L. (1998). The hypoxia-responsive transcription factor EPAS1 is essential for catecholamine homeostasis and protection against heart failure during embryonic development. *Genes Dev* 12, 3320–3324.
- Tian, Y., Yeoh, K., Lee, M., Eriksson, T., Kessler, B.M., Holger, H.B., Edelmann, M.J., William, C., Pugh, C.W., Schofield, C.J., and Ratcliffe, P.J. (2011). Differential sensitivity of hypoxia inducible factor hydroxylation sites to hypoxia and hydroxylase inhibitors. *J Biol Chem.* 286, 13041–13051.
- Vadnais, C., Awan, A. a, Harada, R., Clermont, P.-L., Leduy, L., Bérubé, G., and Nepveu, A. (2013). Long-range transcriptional regulation by the p110 CUX1 homeodomain protein on the ENCODE array. *BMC Genomics* 14, 258.
- Vermeulen, M., Mulder, K.W., Denissov, S., Pijnappel, W.W.M.P., van Schaik, F.M.A., Varier, R.A., Baltissen, M.P.A., Stunnenberg, H.G., Mann, M., and Timmers, H.T.M. (2007). Selective Anchoring of TFIIID to Nucleosomes by Trimethylation of Histone H3 Lysine 4. *Cell* 131, 58– 69.
- Villar, D., Ortiz-Barahona, A., Gomez-Maldonado, L., Pescador, N., Sanchez-Cabo, F., Hackl, H., Rodriguez, B.A.T., Trajanoski, Z., Dopazo, A., Huang, T.H.M., et al. (2012). Cooperativity of Stress-Responsive Transcription Factors in Core Hypoxia-Inducible Factor Binding Regions. *PLoS One* 7.
- Wallace, E.M., Rizzi, J.P., Han, G., Wehn, P. M., Cao, Z., Du, X., Cheng, T., Czerwinski, R.M., Dixon, D.D. Goggin, B.S., Grina, J.S., Halfman, M.M., Maddie, M.A., Olive, S.R., Schlachter, S.T., Tan, Huiling., Wang, B., Wang, K., Xie, Z., Xu, R., Yang, H., Josey, J.A. (2016). A small-molecule antagonist of HIF2α is efficacious in preclinical models of renal cell carcinoma. *Cancer Res.* 76, 5491-5500.
- Wang, G.L., and Semenza, G.L. (1993). Characterization of hypoxia-inducible factor 1 and regulation of DNA binding activity by hypoxia. *J Biol Chem* 268, 21513–21518.
- Wang, G.L., and Semenza, G.L. (1995). Purification and characterization of hypoxia-inducible factor 1. *J Biol Chem* 270, 1230–1237.

- Wang, G.L., Jiang, B.H., Rue, E.A., and Semenza, G.L. (1995). HYPOXIA-INDUCIBLE FACTOR-1 IS A BASIC-HELIX-LOOP-HELIX-PAS HETERODIMER REGULATED BY CELLULAR O₂ TENSION. *Proc Natl Acad Sci U S A* 92, 5510–5514.
- Wang, V., Davis, D.A., Haque, M., Huang, L.E., and Yarchoan, R. (2005). Differential gene up-regulation by hypoxia-inducible factor-1 α and hypoxia-inducible factor-2 α in HEK293T cells. *Cancer Res.* 65, 3299–3306.
- Wang, Y., Wan, C., Deng, L., Liu, X., Cao, X., Gilbert, S.R., Bouxsein, M.L., Faugere, M.C., Guldberg, R.E., Gerstenfeld, L.C., et al. (2007). The hypoxia-inducible factor α pathway couples angiogenesis to osteogenesis during skeletal development. *J. Clin. Invest.* 117, 1616–1626.
- Warnecke, C., Zaborowska, Z., Kurreck, J., Erdmann, V.A., Frei, U., Wiesener, M., and Eckardt, K.U. (2004). Differentiating the functional role of hypoxia-inducible factor (HIF)-1 α and HIF-2 α (EPAS-1) by the use of RNA interference: erythropoietin is a HIF-2 α target gene in Hep3B and Kelly cells. *Faseb J* 18, 1462–1464.
- Watson, C.J., Collier, P., Tea, I., Neary, R., Watson, J.A., Robinson, C., Phelan, D., Ledwidge, M.T., McDonald, K.M., McCann, A., et al. (2014). Hypoxia-induced epigenetic modifications are associated with cardiac tissue fibrosis and the development of a myofibroblast-like phenotype. *Hum. Mol. Genet.* 23, 2176–2188.
- Wenger, R.H., Kvietikova, I., Rolfs, A., Camenisch, G., and Gassmann, M. (1998). Oxygen-regulated erythropoietin gene expression is dependent on a CpG methylation-free hypoxia-inducible factor-1 DNA-binding site. *Eur. J. Biochem.* 253, 771–777.
- Wenger, R.H., Stiehl, D.P., and Camenisch, G. (2005). Integration of oxygen signaling at the consensus HRE. *Sci. STKE* 2005, re12-re12.
- Wiesener, M.S., Jurgensen, J.S., Rosenberger, C., Scholze, C.K., Horstrup, J.H., Warnecke, C., Mandriota, S., Bechmann, I., Frei, U.A., Pugh, C.W., et al. (2003). Widespread hypoxia-inducible expression of HIF-2 α in distinct cell populations of different organs. *Faseb J* 17, 271–273.
- Williams, C.B., Yeh, E.S., and Soloff, A.C. (2016). Tumor-associated macrophages: unwitting accomplices in breast cancer malignancy. *Npj Breast Cancer* 2, 15025.
- Wu, C., Wong, Y.C., and Elgin, S.C. (1979). The chromatin structure of specific genes: II. Disruption of chromatin structure during gene activity. *Cell* 16, 807–814.
- Wu, D., Potluri, N., Lu, J., Kim, Y., Rastinejad, F. (2015). Structural integration in hypoxia inducible factors. *Nature* 524, 303-308.
- Xia, X., and Kung, A.L. (2009). Preferential binding of HIF-1 to transcriptionally active loci determines cell-type specific response to hypoxia. *Genome Biol* 10, R113.
- Xia, X.B., Lemieux, M.E., Li, W., Carroll, J.S., Brown, M., Liu, X.S., and Kung, A.L. (2009). Integrative analysis of HIF binding and transactivation reveals its role in maintaining histone methylation homeostasis. *Proc Natl Acad Sci U S A* 106, 4260–4265.
- Xiang, L.S., Gilkes, D.M., Hu, H.X., Luo, W.B., Bullen, J.W., Liang, H.J., and Semenza, G.L. (2015). HIF-1 α and TAZ serve as reciprocal co-activators in human breast cancer cells. *Oncotarget* 6, 11768–11778.
- Xiao, Y., Hsiao, T.H., Suresh, U., Chen, H.I.H., Wu, X., Wolf, S.E., and Chen, Y. (2014). A novel significance score for gene selection and ranking. *Bioinformatics* 30, 801–807.
- Yamashita, K., Discher, D.J., Hu, J., Bishopric, N.H., and Webster, K.A. (2001). Molecular regulation of the endothelin-1 gene by hypoxia. Contributions of hypoxia-inducible factor-1, activator protein-1, GATA-2, AND p300/CBP. *J. Biol. Chem.* 276, 12645–12653.
- Yang, X.J., and Seto, E. (2003). Collaborative spirit of histone deacetylases in

- regulating chromatin structure and gene expression. *Curr. Opin. Genet. Dev.* 13, 143–153.
- Yu, B., Miao, Z.H., Jiang, Y., Li, M.H., Yang, N., Li, T., Ding, J. (2009). c-Jun Protects Hypoxia-Inducible Factor-1 α from Degradation via Its Oxygen-Dependent Degradation Domain in a Nontranscriptional Manner. *Cancer Res* 69, 7704–7712.
- Yuen, R.K.C., Chen, B., Blair, J.D., Robinson, W.P., and Michael Nelson, D. (2013). Hypoxia alters the epigenetic profile in cultured human placental trophoblasts. *Epigenetics* 8, 192–202.
- Yun, Z., Maecker, H.L., Johnson, R.S., and Giaccia, A.J. (2002). Inhibition of PPAR γ 2 gene expression by the HIF-1-regulated gene DEC1/Stra13: A mechanism for regulation of adipogenesis by hypoxia. *Dev. Cell* 2, 331–341.
- Zanconato, F., Forcato, M., Battilana, G., Azzolin, L., Quaranta, E., Bodega, B., Rosato, A., Bicciato, S., Cordenonsi, M., and Piccolo, S. (2015). Genome-wide association between YAP/TAZ/TEAD and AP-1 at enhancers drives oncogenic growth. *Nat Cell Biol* 17, 1218–+.
- Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. *Genes Dev* 25, 2227–2241.
- Zeng, M., Shen, J., Liu, Y., Lu, L.Y., Ding, K., Fortmann, S.D., Khan, M., Wang, J., Hackett, S.F., Semenza, G.L., Campochiaro, P.A. (2017). The HIF-1 antagonist acriflavine: visualization in retina and suppression of ocular neovascularization. *Journal of Molecular Medicine*. 95, 417–429.
- Zhang, H., Gao, P., Fukuda, R., Kumar, G., Krishnamachary, B., Zeller, K.I., Dang, C. V., and Semenza, G.L. (2007). HIF-1 inhibits mitochondrial biogenesis and cellular respiration in VHL-deficient renal cell carcinoma by repression of C-MYC activity. *Cancer Cell* 11, 407–420.
- Zhang, H., Bosch-Marce, M., Shimoda, L.A., Yee, S.T., Jin, H.B., Wesley, J.B., Gonzalez, F.J., and Semenza, G.L. (2008a). Mitochondrial autophagy is an HIF-1-dependent adaptive metabolic response to hypoxia. *J. Biol. Chem.* 283, 10892–10903.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008b). Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9, R137.
- Zhang, Y.I., Doyle, K., and Bina, M. (1992). Interactions of HTF4 with E-box motifs in the long terminal repeat of human immunodeficiency virus type 1. *J. Virol.* 66, 5631–5634.
- Zhong, H., De Marzo, A.M., Laughner, E., Lim, M., Hilton, D.A., Zagzag, D., Buechler, P., Isaacs, W.B., Semenza, G.L., and Simons, J.W. (1999). Overexpression of hypoxia-inducible factor 1 α in common human cancers and their metastases. *Cancer Res* 59, 5830–5835.

Chapter 10: Declaration

Except for where acknowledgements have been made, the work outlined in this thesis was performed by the author in the CCMP and NDM Research Buildings at the University of Oxford. The work presented in this thesis has not been submitted for any other degree in this university, or any other education institute.

Chapter 11: Acknowledgements

I am immensely grateful to Prof. Sir Peter Ratcliffe for providing me the opportunity to undertake my DPhil studies in his laboratory. Furthermore, I am extremely grateful to both of my co-supervisors, Peter and Prof. David Mole, for all-round outstanding supervision, support, guidance, and regular scientific debate that has been afforded to me during my DPhil. The four years under Peter and David's supervision has been pivotal to my academic and personal development to-date. I would also like to thank all Ratcliffe Group members (both past and present) for their scientific support during my studies. Although, special thanks must be afforded to Prof. Chris Pugh, Dr. Norma Masson, Dr. Matthew Cockman, Dr. James Platt, Dr. Rafik Salama, Martin Attwood and Catherine King: all of whom never hesitated to sacrifice their own time in order field my questions, help interpret my data, provide technical support, or help organise conference trips.

From a social perspective, many thanks to Matt, Rafik and James P for helping me enhance my table tennis skills to an elite level. Thanks to Martin and James P for regularly listening to my beer-fuelled rants on Friday evenings after lab. And thanks to the rest of the lab's (and the neighbouring Nijman lab) social cohort for enjoying regular outings with me – particularly, Cathrine Nordgaard.

I wish to thank my Transfer and Confirmation of Status Assessors, Dr. Sarah De Val, Prof. Skirmantas Kriaucionis and Assoc. Prof. Ester Hammond for stimulating discussion and outstanding advice during my viva assessments. I am grateful for the professional, personal and academic guidance received from my College Advisor, Assoc. Prof Philippa Hulley. I also wish to thank Prof. Renato Iozzo for being my long-time mentor and for first introducing me to experimental biology in his lab.

Extra special thanks goes to all my friends outside of the laboratory that have filled my four years in Oxford with tremendous amounts of fun and enjoyment.

I dedicate this work to my wonderfully supportive, kind and loving Mother.